sciences & méthodes

HARVEY J. MOTULSKY

Biostatistique

Une approche intuitive

Traduction de la 1re édition anglaise
par Michèle Dramaix-Wilmet et Fabienne Nackers

Révision scientifique d'Annie Robert



Ouvrage original :

Intuitive Biostatistics by Harvey J. Motulsky

This translation of *Intuitive Biostatistics* originally published in English in 1995 is published by arrangement with Oxford University Press.
Cette traduction d'*Intuitive Biostatistics* parue en anglais en 1995 est publiée avec l'autorisation d'Oxford University Press.

Pour toute information sur notre fonds et les nouveautés dans votre domaine de spécialisation, consultez notre site web : http://www.deboeck.be

---

1re édition

Dépôt légal :
Bibliothèque Nationale, Paris : mars 2002
Bibliothèque Royale Albert Ier, Bruxelles : 2002/0074/21

ISBN 2-7445-0092-5
ISSN 1376-2281

Imprimé en Espagne

- Les exemples proviennent de la littérature clinique. À quelques endroits, des exemples avec des données fictives ont été introduits, mais la plupart des exemples proviennent de la littérature médicale récente[1]. De manière à centrer la discussion sur la compréhension statistique de base, les résultats ont parfois été quelque peu simplifiés, sans (il faut l'espérer) masquer l'essentiel de ces résultats.
- Explication de l'approche bayésienne. L'approche bayésienne est une aide pour interpréter les P-valeurs, pour combiner les résultats de tests diagnostiques, pour évaluer un conseil génétique et pour analyser des liaisons génétiques. Alors que la plupart des ouvrages d'introduction ignorent l'analyse bayésienne, celle-ci est abordée de manière un peu détaillée dans la partie IV.

2 MATIÈRE COUVERTE

Les sujets inclus dans ce livre ont été choisis de manière à ce que la matière couverte soit étendue plutôt que vue en profondeur. Ce choix a été effectué en raison du grand nombre de méthodes statistiques couramment utilisées dans la littérature biomédicale. En parcourant n'importe quelle revue médicale ou scientifique, on tombe assez rapidement sur l'utilisation d'une technique statistique qui n'est pas mentionnée dans la plupart des livres d'introduction. Pour guider ceux qui lisent ces articles, ce livre contient de nombreux thèmes absents des autres livres : risques relatifs et rapports de cotes, intervalles de prédiction, tests non paramétriques, courbes de survie, comparaisons multiples, designs d'essais cliniques, calcul de la puissance d'un test, régression non linéaire, interprétation des tests diagnostiques (sensibilité, spécificité, etc.). La régression multiple, la régression logistique, le modèle de régression de Cox des risques instantanés proportionnels, les tests de randomisation et lod scores (log de cotes) ont également été brièvement introduits. L'analyse de variance a reçu une place moins importante que de coutume.

3 CHAPITRES À PASSER

Par rapport aux livres de statistique en général, celui-ci est assez court. Cependant, il contient sans doute encore bien plus que ce que les gens souhaitent lire en matière de statistique. Pour juste prendre connaissance des idées principales de la statistique sans trop de détails, il faut lire les chapitres 1 à 5, 10 à 13 et 19.

Ce livre est destiné à quiconque lit des articles dans la littérature biomédicale et pas seulement à ceux qui s'intéressent aux études cliniques. Les chercheurs fondamentalistes pourraient passer les chapitres 6, 9, 20, 21, 32 et 33 qui traitent de sujets qui ne se rencontrent habituellement pas en recherche fondamentale. Les autres chapitres sont appropriés à la fois pour les cliniciens et les chercheurs fondamentalistes.

[1] Ces publications n'ont pas été sélectionnées parce qu'elles étaient particulièrement bonnes ou particulièrement mauvaises. Il s'agit simplement d'un échantillon d'articles de bonnes revues sur lesquels l'auteur de ce livre est « tombé » en cherchant des exemples (aussi, les auteurs ne doivent pas être particulièrement fiers ou gênés de voir leur travail inclus).

La plupart des scientifiques n'ont pas cette chance. Dans de nombreuses branches de la biologie et tout spécialement en recherche clinique, le chercheur se trouve face à une importante variabilité biologique, il n'est pas capable de contrôler toute les variables utiles et est intéressé par de petits effets (disons des changements de l'ordre de 20 %). Avec de telles données, il est difficile de distinguer le signal que l'on recherche du bruit occasionné par la variabilité biologique et les mesures imprécises. Les calculs statistiques sont nécessaires pour extraire quelque chose de sensé de telles données.

3 LES CALCULS STATISTIQUES PERMETTENT D'EXTRAPOLER DE L'ÉCHANTILLON À LA POPULATION

Les calculs statistiques nous permettent de tirer des conclusions générales à partir d'un nombre limité de données. Nous pouvons extrapoler vers un cas plus général à partir de nos données. Les statisticiens disent que l'on extrapole d'un *échantillon* à une *population*. La distinction entre échantillon et population est la clé pour comprendre une grande partie de la statistique. Ci-dessous sont évoqués quatre contextes différents où ces termes « population » et « échantillon » sont utilisés.

- *Contrôle de qualité.* C'est dans le contexte du contrôle de qualité où l'échantillon est sélectionné aléatoirement dans la population générale que les termes *échantillon* et *population* prennent toute leur signification. Par exemple, une usine réalise des lots d'items (la population) mais sélectionne aléatoirement quelques items (l'échantillon) pour les tester. Les résultats dérivés de l'échantillon sont utilisés afin d'effectuer des inférences pour l'entièreté de la population.
- *Sondages politiques.* Un échantillon aléatoire d'électeurs (l'échantillon) est sondé et les résultats sont utilisés pour tirer des conclusions à propos de la population totale d'électeurs.
- *Études cliniques.* L'échantillon de patients étudiés est rarement un échantillon extrait aléatoirement d'une population plus grande. Cependant les patients inclus dans l'étude sont représentatifs d'autres patients semblables et l'extrapolation de l'échantillon à la population a toujours son utilité. Les avis divergent souvent au sujet de la définition précise de la population. La population est-elle constituée de tous les patients semblables qui fréquentent un centre médical particulier ou de tous ceux qui fréquentent l'hôpital universitaire d'une grande ville ou de tous les patients semblables du pays, du monde ? Même si la population est définie de façon plutôt vague, il n'en reste pas moins vrai que l'on désire utiliser les données de l'échantillon afin de tirer des conclusions pour un groupe plus grand.
- *Expériences de laboratoire.* Étendre les notions d'*échantillon* et *population* pour les utiliser dans le cadre des expériences de laboratoire n'est pas chose facile. Les données de l'expérience réalisée constituent l'échantillon. Si l'on répétait l'expérience, on aurait un échantillon différent. Les données de toutes les expériences que l'on aurait pu réaliser constituent la population. À partir des données de l'échantillon, on veut effectuer des inférences pour la situation idéale.

6 POURQUOI EST-CE AUSSI DIFFICILE D'APPRENDRE LA STATISTIQUE ?

Cinq facteurs rendent l'apprentissage de la statistique difficile pour beaucoup d'étudiants :

- La terminologie est trompeuse. En statistique, une signification particulière est attribuée à pas mal de mots ordinaires. Pour comprendre la statistique, il faut comprendre que le sens statistique de termes tels que *significatif*, *erreur* et *hypothèse*, est différent de celui que l'usage courant donne habituellement à ces mots. En lisant ce livre, il faut faire particulièrement attention aux termes statistiques qui correspondent à des mots bien connus.
- Beaucoup de gens semblent croire que les calculs statistiques sont magiques et peuvent conduire à des conclusions beaucoup plus solides que ce n'est le cas réellement. L'expression *statistiquement significatif* est séduisante et souvent mal interprétée.
- La statistique exige la maîtrise de concepts abstraits. Il n'est pas simple de réfléchir à des concepts théoriques tels que populations, distributions de probabilité et hypothèse nulle.
- La statistique est à l'interface entre la mathématique et les sciences. Pour vraiment saisir les concepts statistiques, il faut être capable de les aborder sous les deux angles. Ce livre souligne le côté scientifique et évite la mathématique. Celui qui réfléchit comme un mathématicien peut plutôt préférer un texte qui utilise une approche mathématique.
- La dérivation de nombreux tests statistiques inclut des mathématiques assez compliquées. À moins d'étudier des textes plus approfondis, il faut accepter en confiance, une grande partie des statistiques. On peut très bien apprendre à utiliser des tests statistiques et à en interpréter les résultats même si on ne comprend pas entièrement leur mécanisme. Cette situation est fréquente en science ; en effet, peu de scientifiques comprennent véritablement tous les outils qu'ils utilisent. On peut interpréter les résultats d'un pH mètre (mesure l'acidité) ou d'un compteur à scintillation (mesure la radioactivité), même si l'on ne comprend pas à fond leur mécanisme. Il faut seulement en savoir assez sur le fonctionnement des instruments de mesure pour éviter de les utiliser dans des situations non appropriées. De la même manière, pour autant que l'on utilise les tests statistiques de façon adéquate, on peut calculer des tests statistiques et interpréter leurs résultats même si l'on ne comprend pas comment les équations ont été établies.

2

L'INTERVALLE DE CONFIANCE D'UNE PROPORTION

1 PROPORTIONS VERSUS MESURES

Les résultats d'une expérience peuvent être exprimés de différentes manières. Dans ce chapitre, nous allons considérer uniquement les résultats exprimés sous forme de proportion ou de fraction. Par exemple : la proportion de patients infectés à la suite d'une intervention, la proportion de patients atteints d'infarctus du myocarde chez qui survient un arrêt cardiaque, la proportion d'étudiants qui réussissent un examen, la proportion d'électeurs qui votent pour un certain candidat. Plus loin, nous discuterons d'autres types de variables, notamment des mesures et des temps de survie.

2 LA DISTRIBUTION BINOMIALE : DE LA POPULATION À L'ÉCHANTILLON

Si on lance une pièce de monnaie non truquée, il y a une probabilité (ou chance) de 50 % qu'elle atterrisse sur « face » et une probabilité de 50 % qu'elle atterrisse sur « pile ». Ceci signifie qu'à long terme, une pièce tombera sur face à peu près aussi souvent qu'elle tombera sur pile. Cependant, pour une série particulière de jets de la pièce de monnaie, on peut très bien ne pas observer que la pièce retombe sur face exactement dans la moitié des cas. On peut très bien observer uniquement des « faces » ou uniquement des « piles ».

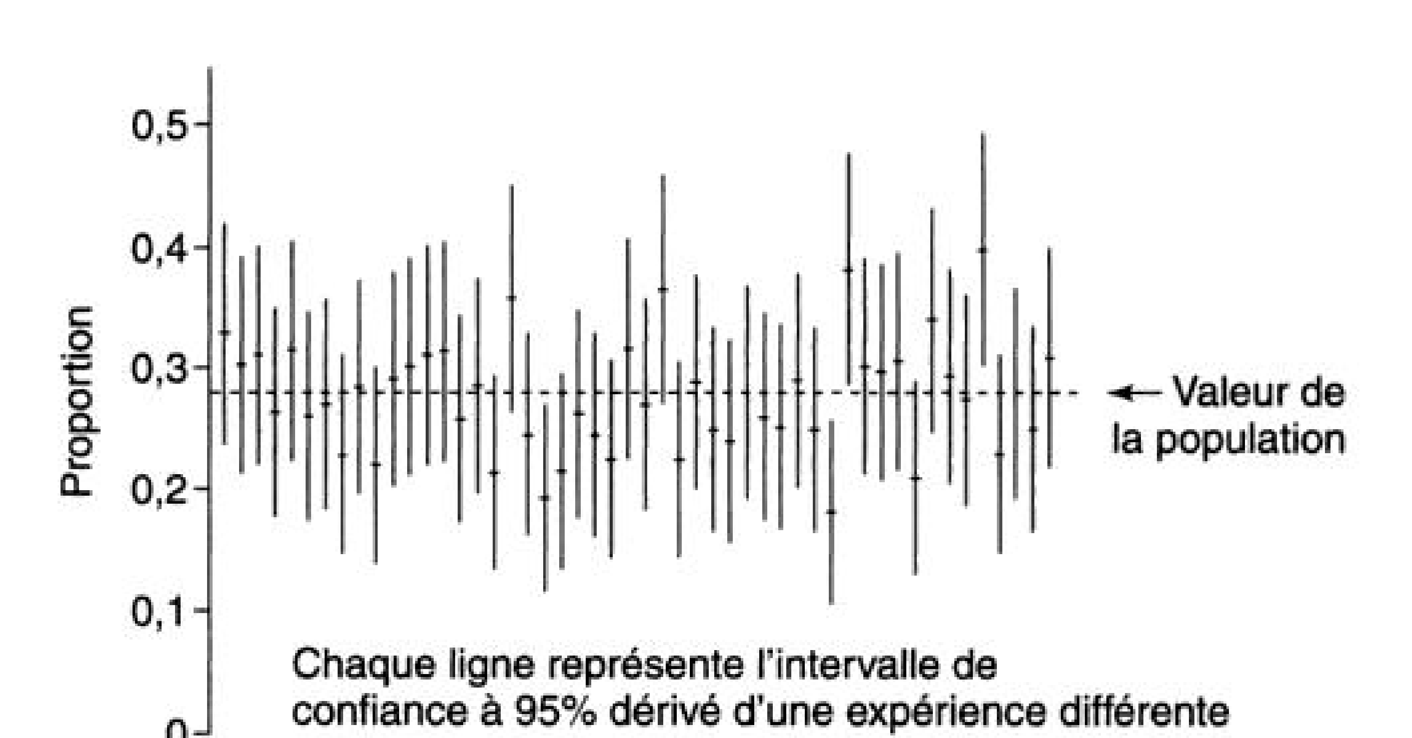


FIGURE 2.1 Interprétation d'un intervalle de confiance. Dans cet exemple, on connaît la proportion vraie de « succès » dans la population qui vaut 0,28. Celle-ci est représentée par la ligne horizontale en pointillés. En utilisant un programme informatique qui peut générer des nombres aléatoires, 50 échantillons ont été sélectionnés de manière aléatoire simple de la population. Chaque ligne verticale représente l'IC à 95 % de la proportion de succès, calculé à partir d'un seul de ces 50 échantillons. Pour la plupart des échantillons, la proportion observée est proche de la proportion vraie, mais pour certains échantillons, l'écart est grand. Pour quatre échantillons (par exemple le cinquième à partir de la droite), l'IC à 95 % n'inclut pas la valeur vraie. Si l'on collectait les données d'un très grand nombre d'échantillons, on s'attendrait à ce type d'observation dans 5 % des cas.

population. Il n'y a pas de moyen de savoir si l'IC à 95 % que l'on a calculé inclut la valeur de la population. Tout ce que l'on sait, c'est qu'à longue échéance, 95 % des intervalles de ce type contiendront la valeur de la population et 5 % ne la contiendront pas. Bien entendu, chaque IC calculé contiendra la proportion observée dans l'échantillon. Ce que l'on ne peut savoir avec certitude, c'est si l'intervalle contient aussi la proportion de la population.

Remarquons que l'IC n'est pas toujours symétrique autour de la proportion observée dans l'échantillon. Pour le premier exemple, il s'étire plus loin vers la droite que vers la gauche. Avec des échantillons plus grands, les IC à 95 % deviennent plus étroits et plus symétriques.

5 HYPOTHÈSES REQUISES

L'IC pourra être interprété correctement si les conditions suivantes sont remplies :

- Échantillon aléatoire simple (ou représentatif)
- Observations indépendantes
- Classification correcte
- Estimation de l'événement réellement intéressant

8 EXEMPLE

Dans le but de mieux conseiller les parents d'enfants prématurés, M.C. Allen et coll. ont analysé la survie de prématurés[3]. Ils ont étudié de manière rétrospective tous les enfants prématurés nés entre 22 et 25 semaines de gestation à l'hôpital Johns Hopkins durant une période de 3 ans. Les chercheurs ont classé les décès en fonction de l'âge gestationnel. Aucun des 29 enfants nés à 22 semaines de gestation n'a survécu jusqu'à 6 mois. Parmi les 39 enfants nés à 25 semaines de gestation, 31 ont survécu au moins 6 mois.

Les auteurs de la recherche ont présenté ces résultats sans IC mais on peut les calculer. Calculer un IC a un sens uniquement si l'échantillon est représentatif d'une population plus grande, pour laquelle on souhaite faire des inférences. Il est raisonnable de penser que ces données récoltées sur plusieurs années dans un hôpital sont représentatives des données d'autres années provenant d'autres hôpitaux, au moins des hôpitaux universitaires des grandes villes aux États-Unis. Si l'on ne veut pas admettre cette hypothèse, il ne faut pas calculer un IC. Cependant, les données ne vaudraient pas la peine d'être récoltées si les chercheurs ne pensaient pas que des résultats similaires pourraient être observés dans le futur, dans d'autres hôpitaux.

Pour les enfants nés à 25 semaines de gestation, on veut déterminer l'IC à 95 % sur base de 31/39. Ces valeurs ne sont pas reprises dans la table A5.1, il faut donc calculer l'IC avec un ordinateur ou à la main. Avec le programme « InStat », on obtient un IC à 95 % qui va de 63 % à 91 % (si l'on utilise l'équation 2.1 qui suit, on calcule un intervalle approximatif : 67 % à 92 %). Ceci signifie que si la proportion réelle d'enfants survivants valait moins de 63 %, il y aurait moins de 2,5 % de chance d'observer par hasard une proportion aussi grande que celle observée dans l'échantillon. Cela signifie aussi que si la proportion réelle était supérieure à 91 %, la chance d'observer juste par hasard, une proportion aussi petite, est inférieure à 2,5 %. Il nous reste donc 95 % de chance (100% – 2,5% – 2,5%) que la proportion vraie se trouve entre 63 % et 91 %.

Pour les enfants nés à 22 semaines de gestation, on veut déterminer l'IC à partir de 0/29. Il ne faut pas utiliser l'équation 2.1 car le numérateur est trop petit. Il faut utiliser un programme informatique ou la table A5.1 dans les annexes. L'IC à 95 % s'étend de 0 % à 11,9 %. On peut être « certain » à 95 % que dans la population, la proportion d'enfants survivants se situe quelque part dans cet intervalle (comme on a observé 0 %, l'IC en réalité ne s'étend que dans une seule direction, de sorte que l'on peut être certain à 97,5 % plutôt que certain à 95 %). Bien qu'aucun enfant de notre échantillon né à 22 semaines d'âge gestationnel n'ait survécu, notre IC inclut la possibilité qu'il y ait dans la population une proportion de survivants aussi élevée que 11,9 %. Ceci signifie que si la proportion globale de survivants dans la population avait n'importe quelle valeur supérieure à 11,9 %, il y aurait moins de 2,5 % de chance d'observer par hasard 0 survivant dans un échantillon de 29 enfants.

[3] MC Allen, PK Donohue, AE Dusman. The limit of viability-Neonatal outcome of infants born at 22 to 25 weeks gestation. *N Engl J Med* 329:1597-1601, 1993.

OBJECTIFS

1. Les termes suivants doivent vous être familiers :
   - Distribution binomiale
   - Distribution binomiale cumulée
   - Intervalle de confiance
   - Échantillon
   - Population
   - Échantillon aléatoire
   - Indépendance
2. Lorsque vous rencontrez une proportion dans un article scientifique (ou ailleurs), vous devriez être capables de :
   - Définir la population.
   - Calculer (ou déterminer à partir d'une table) l'IC à 95 % de la proportion.
   - Interpréter l'IC à 95 % dans le contexte de l'étude.
   - Spécifier les hypothèses à vérifier pour que l'interprétation soit valide.

EXERCICES

1. Parmi les 100 premières personnes qui ont subi une intervention chirurgicale, 6 sont décédées. Pouvez-vous calculer l'IC à 95 % de la probabilité de décéder suite à cette intervention? Si oui, calculez cet intervalle. Si non, de quelles informations devez-vous disposer ? Quelles hypothèses devez-vous faire ?
2. Un nouveau médicament est testé chez 100 patients et diminue la pression artérielle en moyenne de 6 %. Pouvez-vous calculer l'IC à 95 % de la réduction relative de la pression artérielle par ce médicament? Si oui, calculez cet intervalle. Si non, de quelles informations devez-vous disposer ? En quoi consiste l'IC de la réduction relative de la pression artérielle ? Quelles hypothèses devez-vous faire ?
3. On étudie la viabilité de cellules par une coloration au bleu de trypan. Avec un hémocytomètre, on dénombre 94 cellules non teintées (viables) et 6 cellules teintées (indiquant qu'elles ne sont pas viables). Pouvez-vous calculer l'IC à 95 % de la proportion de cellules teintées (mortes) ? Si oui, calculez cet intervalle. Si non, de quelles informations devez-vous disposer ? Quelles hypothèses devez-vous faire ?
4. En 1989, 20 étudiants sur 125 inscrits en seconde année de médecine à San Diego ont raté l'examen écrit de biostatistique. Pouvez-vous calculer l'IC à 95 % de la probabilité de réussir cet examen ? Si oui, calculez cet intervalle. Si non, de quelles informations devez-vous disposer ? Quelles hypothèses devez-vous faire ?
5. Ross Perot a obtenu 19 % des votes lors de l'élection présidentielle aux États-Unis en 1992. Pouvez-vous calculer l'IC à 95 % du pourcentage d'électeurs qui ont voté pour lui ? Si oui, calculez cet intervalle. Si non, de quelles informations devez-vous disposer ? Quelles hypothèses devez-vous faire ?

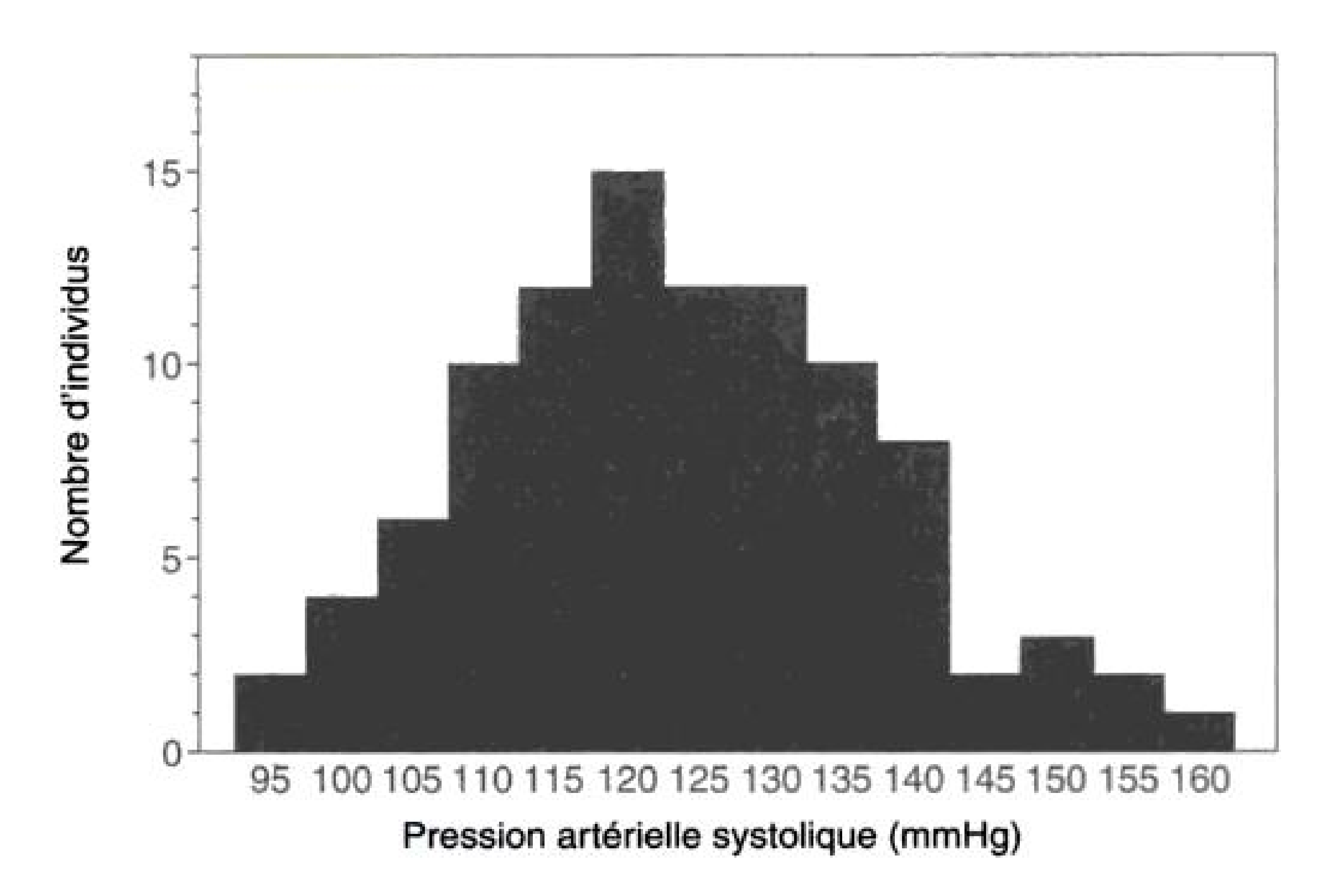


FIGURE 3.1 Histogramme des pressions artérielles systoliques de 100 étudiants. Chaque rectangle couvre un intervalle de 5 mmHg sur l'axe horizontal. La hauteur de chaque rectangle représente le nombre d'étudiants dont la pression artérielle est incluse dans cet intervalle.

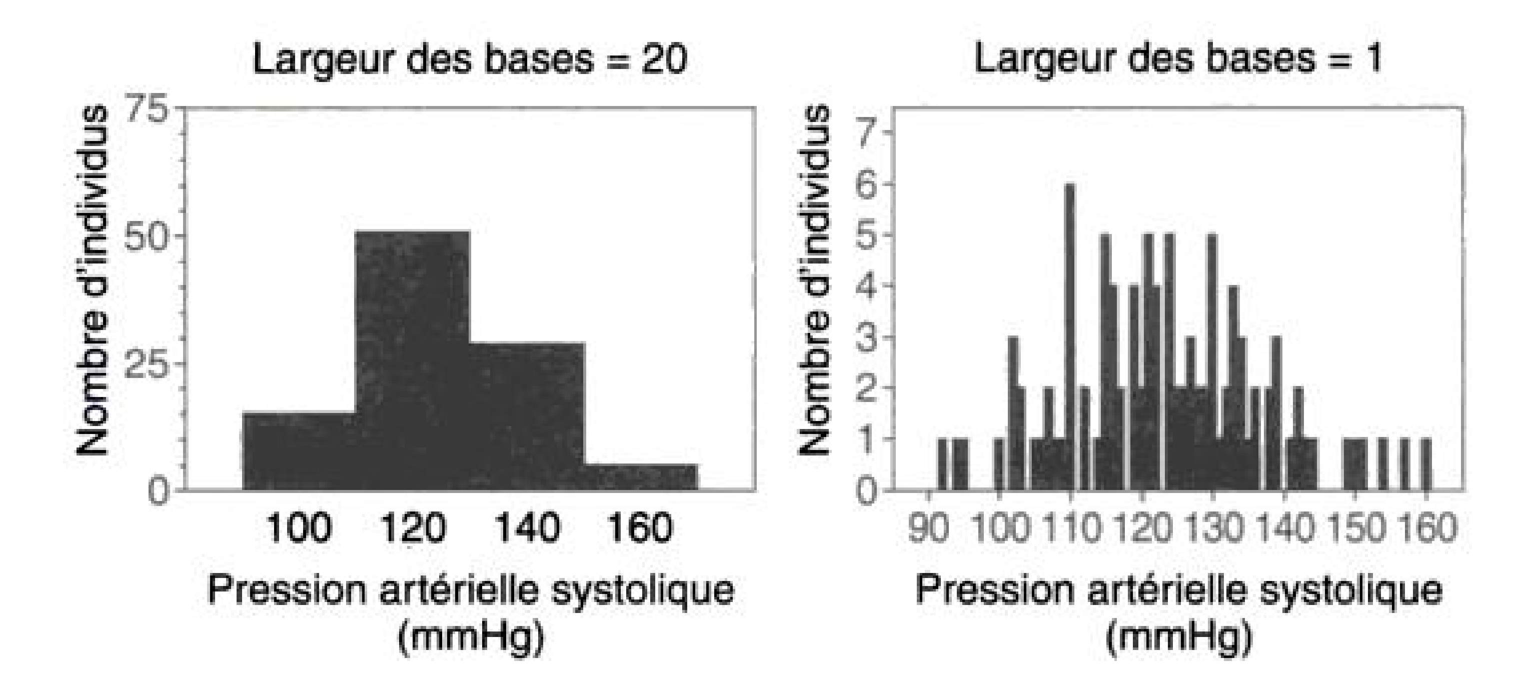


FIGURE 3.2 D'autres histogrammes de la pression artérielle systolique de 100 étudiants. À gauche, chaque rectangle couvre un intervalle de 20 mmHg. Comme il y a très peu de rectangles, le graphique ne fournit pas beaucoup d'informations. À droite, chaque rectangle a une largeur de 1 mmHg. Comme il y a trop de rectangles, le graphique est trop détaillé et il est difficile de voir la distribution des valeurs. Les histogrammes sont généralement les plus utiles lorsqu'ils incluent environ 10 à 20 rectangles.

- Les valeurs enregistrées peuvent différer de manière substantielle de la pression artérielle réelle. Chaque mesure a été effectuée par une personne différente et inexpérimentée. Mesurer la pression artérielle avec un tensiomètre est quelque peu subjectif étant donné que cela requiert de noter la position d'une colonne de mercure à l'instant où un faible son apparaît (systolique) ou disparaît (diastolique). Il est également important de placer le brassard correctement. Il faut de l'expérience.

6 N OU n – 1 ? LA DÉVIATION STANDARD DE LA POPULATION VERSUS LA DS DE L'ÉCHANTILLON

Les équations 3.1 et 3.2 supposent que l'on a effectué des mesures sur l'entièreté de la population. Comme nous l'avons déjà vu, c'est rarement le cas. Tout l'intérêt des calculs statistiques est d'effectuer des inférences pour la population entière à partir des mesures effectuées dans un échantillon. Ceci introduit une complication supplémentaire dans le calcul de la variance et la DS. Pour calculer la DS à partir de l'équation 3.2, il faut calculer l'écart de chaque valeur par rapport à la moyenne de la population. Mais on ne connaît pas la moyenne de la population. Tout ce que l'on connaît, c'est la moyenne de l'échantillon. Sauf dans les cas très rares où la moyenne de l'échantillon serait égale à la moyenne de la population, les valeurs sont toujours plus proches (en moyenne) de leur moyenne d'échantillon que de la moyenne de la population globale. La somme des carrés des écarts par rapport à la moyenne de l'échantillon est donc plus petite que la somme des carrés d'écarts par rapport à la moyenne de la population et l'équation 3.2 donne une valeur trop petite pour la DS. Ce problème est résolu en réduisant le dénominateur à n – 1, plutôt que N. Le calcul de la variance et de la DS à partir d'un échantillon se fait en utilisant l'équation 3.3.

$$\text{Variance de l'échantillon} = s^2 = \frac{\sum_{i=1}^{n}(Y_i - m)^2}{n-1}$$
$$\text{DS de l'échantillon} = s = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - m)^2}{n-1}} \qquad (3.3)$$

Remarquons que l'on passe de l'alphabet grec (μ, σ) à l'alphabet romain (m, s) lorsque l'on passe des discussions sur la moyenne et DS de la population aux discussions sur la moyenne et la DS de l'échantillon. Ce livre passe sous silence une foule de détails mathématiques et parfois passe sur cette distinction. Dans des livres plus mathématiques, il faut faire attention à la différence entre lettres grecques et romaines.

> If the difference between N and n – 1 ever matters you, then you are probably up to no good anyway — e.g. trying to substantiate a questionable hypothesis with marginal data.
>
> W.H. Press *et al., Numerical Recipes*

Si la différence entre N et n – 1 vous préoccupe toujours, alors vous faites sans doute fausse route de tout côté — par ex., en essayant de rendre substantielle une hypothèse douteuse avec des données marginales.

Il y a une autre façon de comprendre pourquoi le dénominateur est n – 1 plutôt que N. Lorsque nous calculons la moyenne m de l'échantillon, nous faisons la somme de toutes les valeurs Y et la divisons par le nombre de valeurs, n. Pourquoi diviser par n ? Nous avons appris à calculer une moyenne depuis si longtemps que nous n'y avons probablement jamais songé. Techniquement, la moyenne est définie par la somme divisée par des degrés de liberté. La moyenne de l'échantillon a n degrés de liberté parce que chaque

4

LA DISTRIBUTION GAUSSIENNE

1 LES DISTRIBUTIONS DE PROBABILITÉ

Comme nous l'avons vu dans le chapitre précédent, un histogramme représente graphiquement la distribution des valeurs d'un échantillon. Dans de nombreuses situations, l'histogramme a la forme d'une cloche. Les deux premiers graphiques de la figure 4.1 représentent les distributions en forme de cloche de deux échantillons. Le deuxième échantillon est plus grand et l'histogramme a été construit avec des rectangles plus étroits, de sorte qu'il apparaît plus lisse.

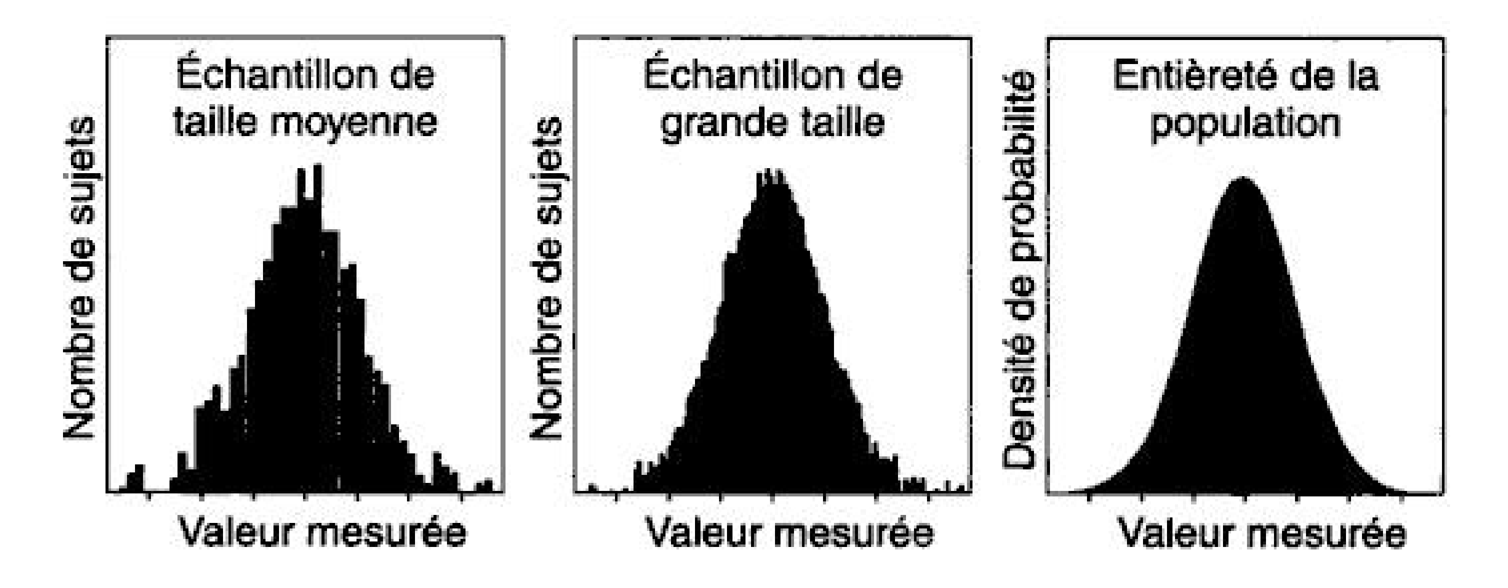


FIGURE 4.1 Histogrammes et distributions de probabilité. Le graphique de gauche représente l'histogramme d'un échantillon de taille moyenne. La hauteur de chaque rectangle représente le nombre de sujets dont la valeur est contenue dans l'intervalle couvert par sa base. Le graphique du milieu représente l'histogramme d'un échantillon plus grand. Les rectangles sont plus étroits. Le graphique de droite représente la distribution de l'entièreté de la population. Comme la population est infinie et qu'il n'y a pas de rectangle, il n'est pas possible que l'axe des Y représente le nombre de sujets. Au lieu de cela, il représente la *densité de probabilité*.

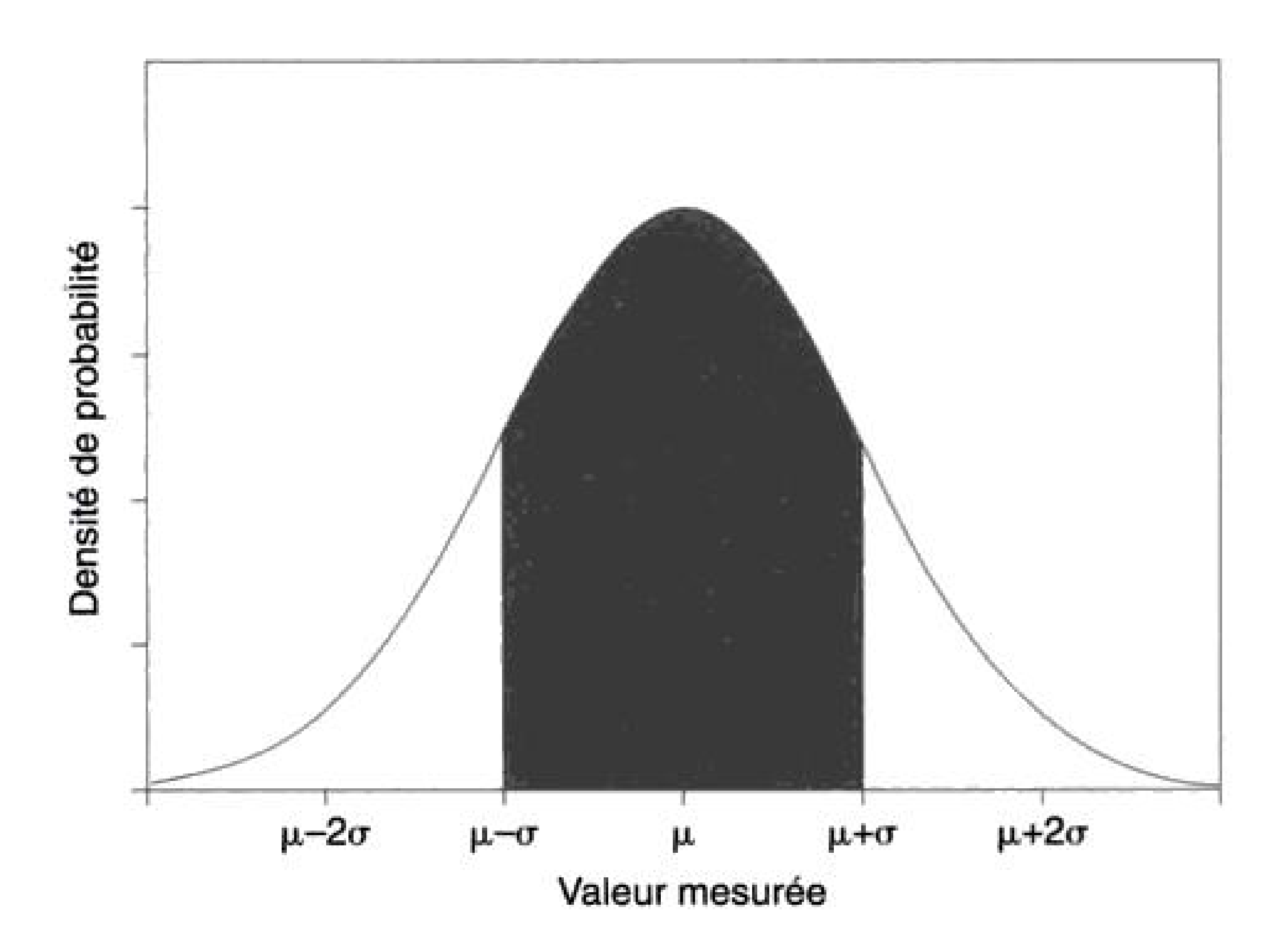


FIGURE 4.3 Interprétation des distributions de probabilité. La surface sous l'entièreté de la courbe représente l'entièreté de la population. Toutes les valeurs comprises entre une DS au-dessus et une DS en dessous de la moyenne sont noircies. Le rapport de la surface noircie sur l'entièreté de la surface correspond à la proportion de la population dont la valeur se trouve entre une DS au-dessus et une DS en dessous de la moyenne. On peut constater que la surface noircie vaut environ deux tiers de la surface totale. Si nous la mesurons avec plus de précision, nous trouvons qu'elle vaut 68,27 % de la totalité de la courbe.

Dans les articles scientifiques, on trouve souvent des moyennes et des DS sans représentation graphique de la distribution des valeurs. En lisant « moyenne = 123, DS = 14 », le lecteur devrait avoir une idée intuitive de ce que cela représente si la population est gaussienne. Les personnes qui ont un hémisphère cérébral droit bien développé devraient être capables de visualiser mentalement la figure 4.4. Celles dont l'hémisphère gauche est dominant devraient être capables d'effectuer mentalement quelques calculs laborieux (deux tiers des valeurs sont compris entre 109 et 137). Dans tous les cas, le lecteur devrait être capable de se représenter la distribution des données sur base de la DS.

La table z permet de répondre à des questions comme celle-ci : quelle proportion de la population a une pression artérielle systolique supérieure à 140 mmHg ? Étant donné que la moyenne vaut 123,4 mmHg et la DS = 14,0, la question porte sur des écarts de plus de $(140{,}0 - 123{,}4)/14{,}0 = 1{,}2$ DS à la moyenne. Considérant $z = 1,2$, la table A5.2 des annexes permet de dire que 11,51 % de la population ont une pression artérielle systolique supérieure à 140 mmHg.

De façon plus générale, il est possible de calculer z pour n'importe quelle valeur de Y en utilisant l'équation 4.1 :

$$z = \frac{|Y - \text{moyenne}|}{DS} \qquad (4.1)$$

plus marqués dans les extrémités de la distribution. Il n'y a pas de raison non plus de penser que la population est distribuée symétriquement de part et d'autre de la moyenne.

- Pourquoi dire que 2,5 % des pressions artérielles dans la population sont anormalement élevées et que 2,5 % des pressions artérielles sont anormalement basses ? Qu'est-ce que ces 2,5 % ont de tellement spécial ? Pourquoi devrait-il y avoir exactement autant de valeurs anormalement élevées que de valeurs anormalement basses ?

Le problème est que le terme *normal* a au moins trois significations :

- Les mathématiciens et les statisticiens utilisent ce terme comme synonyme de l'adjectif gaussien. Ils ne trouvent cependant rien d'« anormal » aux variables distribuées d'une toute autre façon.
- Les scientifiques utilisent habituellement ce terme pour caractériser des valeurs couramment observées.
- Les cliniciens utilisent le mot *normal* tantôt pour dire « habituel », tantôt pour dire que des valeurs ne sont pas associées à la présence d'une maladie.

Déterminer une « étendue normale » est un problème difficile. Tout peut donc sembler un peu confus en ce moment. L'important est de réaliser qu'il est habituellement trop simple de définir une « étendue normale » de valeurs par la moyenne plus ou moins 2 DS.

> Everybody believes in the normal approximation, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact.
>
> G. Lippman (1845-1921)

> Tout le monde croit en l'approximation normale, les expérimentateurs parce qu'ils croient que c'est un théorème mathématique, les mathématiciens parce qu'ils croient que c'est un fait expérimental.



RÉSUMÉ

De nombreuses variables suivent une distribution gaussienne. On s'y attend lorsque de nombreux facteurs indépendants contribuent à la variabilité sans qu'un des facteurs soit plus important. Les surfaces sous des portions de la distribution gaussienne sont reprises dans une table, celle de la distribution « z ». On appelle *intervalle de prédiction* l'étendue des valeurs dans laquelle on a 95 % de chance de trouver la prochaine valeur qui sera échantillonnée dans la population. Les *limites normales* peuvent être définies de nombreuses façons. La distribution gaussienne n'est pas toujours utile pour définir des valeurs « normales ».

Dans l'exemple 5.2, nous avons tiré un échantillon aléatoire de cinq étudiants de la classe. L'hypothèse est donc respectée. Si nous avions choisi des étudiants volontaires pour participer à l'étude, nous aurions mesuré les pressions artérielles d'étudiants particulièrement intéressés par leur pression artérielle (on leur avait peut-être signalé autrefois qu'elle était élevée). Un tel échantillon n'aurait pas été représentatif de l'entièreté de la population et toute inférence statistique nous aurait probablement induits en erreur.

Dans les études cliniques, une sélection aléatoire des patients dans l'entièreté de la population des patients similaires n'est pas réalisable. Au lieu de cela, les patients sont sélectionnés pour participer à l'étude parce qu'ils se sont trouvés dans la clinique adéquate au moment adéquat. On parle plutôt d'*échantillon de convenance* que d'*échantillon aléatoire*. Pour que les calculs statistiques aient un sens, nous devons considérer que l'échantillon de convenance représente correctement l'entièreté de la population et que les résultats sont semblables à ceux qui auraient été observés si on avait utilisé un véritable échantillon aléatoire.

- La population suit une distribution gaussienne, au moins approximativement. Cette condition n'est pas d'une grande importance si l'échantillon est grand. Dans l'exemple 5.1, l'hypothèse d'une distribution gaussienne importe très peu puisque l'échantillon est constitué de 100 personnes (à moins que la distribution de la population ne soit très étrange). L'exemple 5.2 concerne seulement 5 étudiants. Dans le cas d'un échantillon si petit, l'IC ne peut être interprété que si on suppose que la population générale est approximativement gaussienne.
- Chaque sujet provient de la même population et chacun a été sélectionné indépendamment des autres. En d'autres termes, sélectionner un sujet ne devrait pas modifier la probabilité de sélectionner n'importe quel autre. L'exemple de la pression artérielle ne serait pas valide s'il y avait en réalité moins de 100 étudiants et que certains avaient été investigués 2 fois. Il ne serait pas valide non plus si quelques étudiants étaient enfants de la même famille ou jumeaux (comme la pression artérielle est en partie déterminée par des facteurs génétiques, deux enfants de la même famille ont sans doute des pressions artérielles plus proches que deux personnes sélectionnées aléatoirement).

Dans de nombreuses situations, ces hypothèses ne sont pas strictement respectées. Il peut arriver que les patients d'une étude soient plus homogènes que ne le sont les patients de l'entièreté de la population. Des mesures faites dans un laboratoire peuvent avoir une DS plus petite que des mesures faites dans d'autres laboratoires à d'autres moments. De façon plus générale, il est possible que la population qui nous intéresse vraiment soit plus diversifiée que la population dont sont issues les données. Il se peut également que la population ne soit pas gaussienne. Si ne fusse qu'une hypothèse n'est pas respectée, l'IC sera sans doute trop optimiste (trop étroit). Le véritable IC (tenant compte d'un quelconque écart par rapport aux hypothèses) est sans doute plus large que celui qui a été calculé.

distribution en forme de cloche. Plus on augmente la taille de l'échantillon, plus la distribution des moyennes d'échantillons se rapproche d'une distribution gaussienne parfaite.

La distribution de la population importe peu. La distribution des moyennes échantillonnées sera approximativement gaussienne si la taille de l'échantillon est suffisamment grande. À quelle taille cela correspond-t-il ? La réponse est évidente : « cela dépend » ! Cela dépend de la définition du mot « approximativement » et de la distribution de la population. Même si la distribution de la population est vraiment étrange, un échantillon de 100 valeurs est suffisamment grand pour invoquer le théorème central limite. Si la distribution de la population est approximativement symétrique et unimodale (elle ressemble à une seule montagne et pas à une chaîne de montagnes), on peut invoquer le théorème central limite même si l'échantillon compte seulement une douzaine de valeurs.

Quelle est la déviation (DS) de la distribution de moyennes d'échantillons ? Ce n'est pas la même chose que la DS de la population. Puisque les moyennes des échantillons sont distribuées de façon plus compacte que les valeurs dans la population, on s'attend à ce que la DS de la distribution des moyennes d'échantillons soit plus petite que la DS de la population. On s'attend aussi à ce qu'elle dépende de la DS de la population et de la taille de l'échantillon. Si les valeurs varient beaucoup dans la population (grande DS), les moyennes d'échantillons seront plus dispersées que si la population était très compacte (petite DS). Si on collecte des échantillons plus grands, les moyennes des échantillons seront plus proches les unes des autres et, par conséquent, la DS des moyennes échantillonnées sera plus petite.

Le théorème central limite nous enseigne que la DS des moyennes échantillonnées égale la DS de la population divisée par la racine carrée de la taille de l'échantillon. C'est l'origine du rapport $s/\sqrt{n}$ dans l'équation 5.1.

5 L'ERREUR STANDARD DE LA MOYENNE

L'expression *déviation standard des moyennes d'échantillons* n'est pas très pratique. On a donc attribué un nom plus court à cette valeur : l'*erreur standard de la moyenne*, abrégée ESM. On fait souvent référence à l'ESM en utilisant le terme d'*erreur standard*, omettant le mot *moyenne* mais le sous-entendant. Le terme est plutôt mal approprié, l'erreur standard de la moyenne n'ayant habituellement rien à voir avec des standards ni avec des erreurs.

D'après le théorème central limite, l'ESM se calcule sur base de la DS de la population et de la taille de l'échantillon au moyen de l'équation 5.2.

$$\text{ESM} = \frac{\text{DS}}{\sqrt{n}} \tag{5.2}$$

L'ESM quantifie la précision de la moyenne de l'échantillon. Une petite ESM signifie que la moyenne de l'échantillon est sans doute assez proche de la vraie moyenne de la population. Une grande ESM signifie que la moyenne de l'échantillon est probablement assez éloignée de la vraie moyenne de la population.

Notons bien que l'ESM ne quantifie pas directement la dispersion ou la variabilité dans la population. C'est une erreur fréquente. Une petite ESM peut résulter davantage

- Nous pouvons penser aux sources de dispersion. La variabilité de la PA résulte de nombreuses variables génétiques et environnementales ainsi que d'imprécisions dans les mesures. Lorsque la dispersion résulte d'une somme de nombreux facteurs, on s'attend à observer une distribution gaussienne ou au moins approximativement gaussienne[11].
- Nous pouvons tester si une distribution de données est compatible avec une distribution gaussienne en réalisant des calculs formels. Pour plus d'informations, il faut se renseigner sur le test de Kolmogorov-Smirnov dans un livre plus avancé.

Que doit-on faire si la distribution des données s'écarte de façon importante d'une distribution gaussienne ? Il a trois possibilités.

- On peut transformer mathématiquement les valeurs pour transformer une population non gaussienne en une population gaussienne. Il faut transformer chaque valeur en son logarithme, son inverse ou sa racine carrée (ou encore une autre fonction). Cela peut paraître un peu curieux mais c'est une bonne méthode. Souvent, cette transformation est biologiquement ou chimiquement justifiée. Par exemple, il est souvent sensé, à la fois biologiquement et statistiquement, d'exprimer l'acidité par le pH plutôt que par la concentration d'ions hydrogène, d'exprimer l'efficacité d'un produit pharmaceutique en $\log(CE_{50})$[12] plutôt que par le CE_{50} et d'exprimer la fonction rénale par l'inverse de la concentration de créatinine plasmatique plutôt que par la concentration de créatinine plasmatique elle-même.
- Nous pouvons nous reposer sur le théorème central limite et analyser de grands échantillons en utilisant des méthodes statistiques basées sur la distribution gaussienne même si les populations ne sont pas gaussiennes. Nous pouvons compter sur le théorème central limite si les deux conditions suivantes sont respectées : (1) Nous faisons de l'inférence sur la moyenne de la population et pas sur les détails de la distribution elle-même. (2) Soit les échantillons sont très grands, soit la population est approximativement gaussienne.
- Nous pouvons utiliser des méthodes statistiques qui ne sont pas basées sur la distribution gaussienne. Par exemple, il est possible de calculer l'IC à 95 % d'une médiane sans faire aucune hypothèse sur la distribution de la population. Nous discuterons certaines de ces méthodes, appelées non paramétriques, plus loin dans le livre.

8 L'INTERVALLE DE CONFIANCE D'UNE PROPORTION REVISITÉ[13]

Nous avons déjà rencontré l'équation permettant de calculer l'IC d'une proportion (équation 2.1). Elle est reprise dans l'équation 5.6.

[11] Il faut aussi supposer que tous ces facteurs ont approximativement le même poids.

[12] Le CE_{50} est la concentration d'un produit pharmaceutique nécessaire pour obtenir la moitié de l'effet maximal.

[13] Cette partie est plus avancée que les autres. Elle peut être passée sans perdre le fil de l'exposé.

5. Vous devez être capables de passer d'une DS à une ESM sans l'aide d'un livre.
6. Vous devez être capables d'interpréter les barres d'erreur présentées dans les graphiques ou les tableaux des articles publiés.

EXERCICES

1. La figure 5.4 montre la distribution de probabilité du temps d'ouverture de canaux ioniques. La plupart des canaux restent ouverts pendant un temps très court et quelques-uns restent ouverts plus longtemps. Imaginez que vous mesurez le temps d'ouverture de 10 canaux et que vous calculez le temps moyen. Vous répétez ensuite cette expérience de nombreuses fois. Quelle est la forme attendue de la distribution des moyennes de temps d'ouverture ?

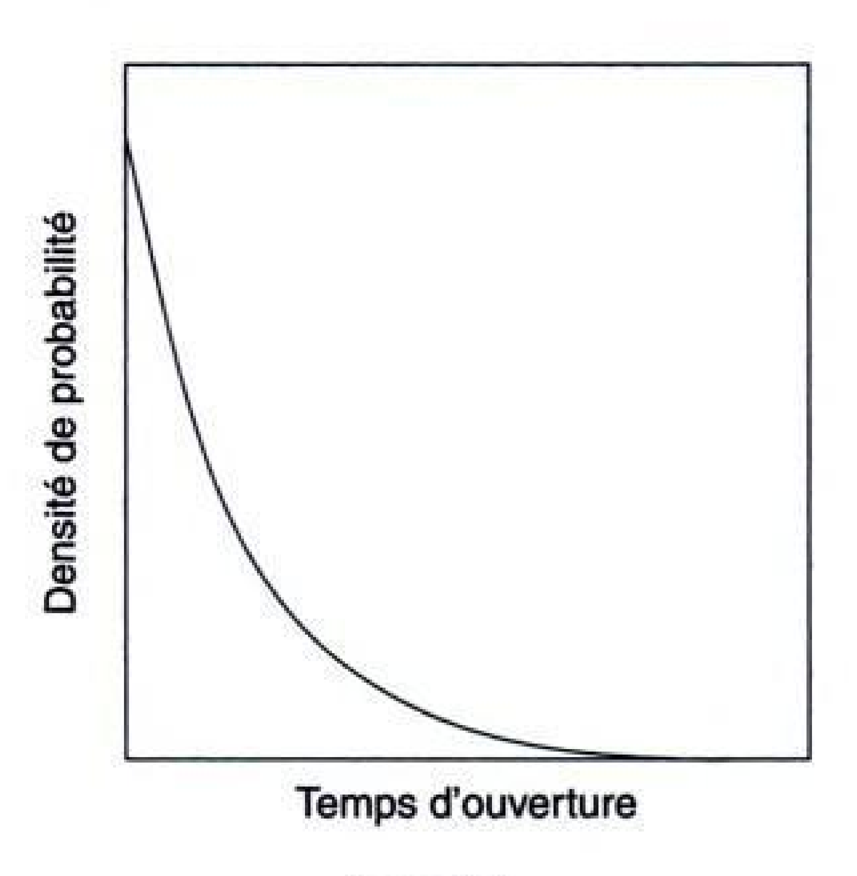


FIGURE 5.4

2. On a mesuré l'activité enzymatique de cellules en culture. L'expérience a été répétée trois jours différents et chaque jour, la mesure a été répétée trois fois dans des conditions expérimentales identiques. Le seul but de la répétition des expériences était de déterminer la valeur avec plus de précision. Les résultats rapportés représentent l'activité enzymatique en unités par minute par milligramme de protéine membranaire.

	Mesure 1	Mesure 2	Mesure 3
Lundi	234	220	229
Mardi	269	967	275
Mercredi	254	249	246

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

2 LES DONNÉES DE SURVIE CENSURÉES

Dans l'exemple précédent, nous savions pour tous les sujets s'ils étaient décédés avant 36 mois ou s'ils avaient survécu plus de 36 mois (extrémité droite de notre courbe). Les données réelles sont rarement aussi simples. Dans la plupart des études de survie, des sujets survivants ne sont pas suivis pendant tout le laps de temps correspondant à l'entièreté de la courbe. Cela peut arriver dans deux situations :

- Quelques personnes sont encore en vie à la fin de l'étude mais n'ont pas été suivies pendant tout le laps de temps couvert par l'entièreté de la courbe. Beaucoup d'études enrôlent des patients pendant une période de plusieurs années. Les patients qui sont enrôlés tardivement ne sont pas suivis autant d'années que les patients qui sont enrôlés plus tôt. Imaginons une étude qui enrôle des patients entre 1986 et 1989 et qui se termine fin 1991. Un patient A peut avoir été enrôlé au début de l'année 1989 et être toujours en vie à la fin de l'étude. Même si l'étude a duré 6 ans, nous savons seulement que le patient A a survécu au moins 3 ans.
- Certains sortent de l'étude précocement. Ils peuvent avoir déménagé dans une autre ville ou s'être lassés des hôpitaux universitaires. Un patient B peut avoir été enrôlé en 1986 mais avoir déménagé dans une autre ville (et avoir cessé de suivre le protocole) en 1988. Nous savons que ce sujet a survécu au moins 2 ans en suivant le protocole mais nous ne pouvons pas évaluer sa survie après cela.

Dans les deux cas, nous savons que le sujet a survécu jusqu'à un certain moment mais nous n'avons aucune information utile sur ce qui s'est passé par la suite. L'information concernant ces patients est dite « censurée ». Avant le moment de censure, nous savons qu'ils étaient en vie et qu'ils suivaient le protocole expérimental ; ces sujets contribuent donc à de l'information intéressante. Dès qu'ils ont été censurés, nous ne pouvons plus utiliser aucune information les concernant. Soit nous n'avons simplement pas d'information au-delà du jour de censure (parce que les données n'ont pas été ou ne pouvaient pas être collectées), soit nous avons l'information mais nous ne pouvons pas l'utiliser (parce que le patient n'a pas suivi plus longtemps le protocole expérimental). Le terme « censure » a une connotation négative. Il semble dire que le sujet a fait quelque chose de mal mais ce n'est pas le cas. Ce sont les données qui ont été censurées, pas le sujet !

3 TRACER UNE COURBE DE SURVIE

Il existe deux méthodes légèrement différentes pour tracer une courbe de survie. Dans la méthode actuarielle, l'axe des X est divisé en intervalles réguliers, par exemple en mois ou en années et la survie est calculée pour chaque intervalle. Dans la méthode de Kaplan-Meier, la survie est recalculée chaque fois qu'un patient décède. Il faut préférer cette méthode, sauf si le nombre de patients est énorme. Le terme *d'analyse de tables de survie* est utilisé de façon inconsistante et inclut d'habitude les deux méthodes. Il vaut mieux reconnaître les trois noms.

La méthode de Kaplan-Meier suit une logique simple mais elle est fastidieuse. Puisque les calculs peuvent être réalisés par des logiciels informatiques, leurs détails ne

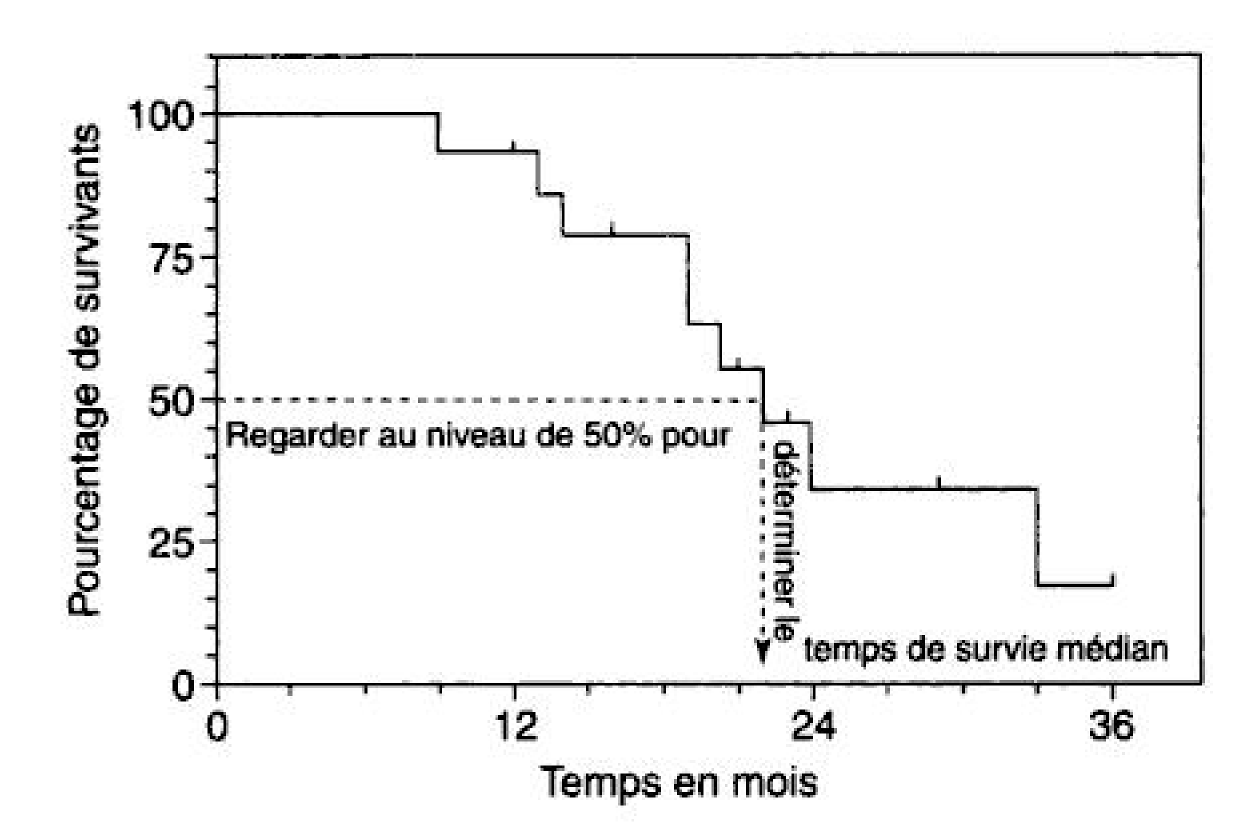


FIGURE 6.4 Survie médiane. La médiane est le percentile 50. Le temps de survie médian est le temps qui s'est écoulé jusqu'à ce que la moitié des sujets soient morts et auquel la moitié des sujets sont encore vivants. Il faut regarder au niveau de 50 % de survivants pour déterminer le temps de survie médian. Si moins de la moitié des sujets sont décédés à la fin de l'étude, il est impossible de déterminer la survie médiane.

- *Observations indépendantes.* La survie de n'importe quel sujet dans la population ne doit pas influencer la probabilité de survie de n'importe quel autre sujet.
- *Critère d'entrée invariable.* Les patients sont inclus dans les études sur une période de plusieurs mois ou de plusieurs années. Il est donc important dans ces études que les critères de départ ne changent pas au cours de la période d'enrôlement. Imaginons une courbe de survie commençant en date du diagnostic de la première métastase d'un cancer. Que se passerait-il si une amélioration de la technique diagnostique permettait de détecter les métastases plus précocement ? Même sans aucun changement dans le traitement ni dans l'histoire naturelle de la maladie, le temps de survie semblera augmenter (les patients décèdent au même âge mais, le diagnostic étant posé à un plus jeune âge, ils vivent plus longtemps avec le diagnostic).
- *Critère invariable pour définir la survie.* Si la courbe représente graphiquement le temps jusqu'au décès, le critère final est assez clair. Si la courbe représente graphiquement le temps jusqu'à un autre événement, il est crucial que l'événement soit évalué de façon identique tout au long de l'étude.
- *Le temps de censure n'est pas lié à la survie.* La survie des patients censurés doit être identique (en moyenne) à la survie de ceux qui ne le sont pas. Si de nombreux patients sont censurés, cette hypothèse est capitale pour garantir la validité des résultats. Il n'y a pas de raison de douter de cette hypothèse pour les patients en vie à la fin de l'étude. Le problème est différent avec les patients qui ont abandonné l'étude. Il faut se demander pourquoi les patients ont quitté l'étude. Si la raison était liée à la survie, la courbe de survie ne serait pas juste. Une courbe de survie, par exemple, induirait en erreur si de nombreux patients étaient sortis de l'étude parce qu'ils étaient trop malades pour se rendre à la clinique ou parce qu'ils se

Jusqu'à présent, nous avons seulement analysé des données provenant d'un seul groupe. D'habitude, quand on analyse des données, on souhaite comparer deux groupes (ou davantage). Cette partie explique comment y parvenir en calculant des intervalles de confiance. Nous apprendrons plus tard comment comparer des groupes en calculant des P-valeurs et en déterminant la signification statistique.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

2 LE CALCUL DE L'IC À 95 % DE LA DIFFÉRENCE ENTRE LES MOYENNES DE GROUPES NON APPARIÉS[17]

Afin de calculer l'IC à 95 % de la différence entre des moyennes, il faut d'abord calculer l'ES de la différence sur base des erreurs standards (ES) de chacun des deux groupes, ESM_a et ESM_b. Si les deux groupes sont de même taille, l'équation 7.1 permet de calculer l'ESM de la différence entre les moyennes des deux groupes :

$$\text{ES de la différence} = \sqrt{ESM_a^2 + ESM_b^2} \qquad (7.1)$$

Si les deux groupes sont de tailles inégales, une équation plus compliquée est nécessaire (7.2) puisque l'ESM du groupe le plus grand doit avoir plus de poids que l'ESM du groupe plus petit. L'équation se construit avec les DS (DS_a et DS_b) et les tailles (n_a et n_b) des deux échantillons :

$$DS_{commune} = \sqrt{\frac{(n_a - 1) \cdot DS_a^2 + (n_b - 1) \cdot DS_b^2}{n_a + n_b - 2}}$$

$$\text{ES de la différence} = DS_{commune}\sqrt{\frac{1}{n_a} + \frac{1}{n_b}} \qquad (7.2)$$

L'IC à 95 % de la différence entre les moyennes (Δ) se calcule au moyen de l'équation 7.3 :

IC à 95 % de la différence entre les moyennes :

$$(\Delta - t^* \cdot \text{ES de la différence}) \text{ à } (\Delta + t^* \cdot \text{ES de la différence})$$

$$dl = n_a + n_b - 2 \qquad (7.3)$$

Si on applique la première équation aux données de l'échantillon, l'ES de la différence entre les deux moyennes vaut 30,0 ml/kg, ce qui est un peu plus grand que l'ESM de chacun des deux groupes. C'est logique puisque l'incertitude qui règne au sujet de la différence entre les moyennes est plus grande que celle qui règne au sujet de chacune des deux moyennes. En d'autres termes, les erreurs s'additionnent de sorte que la différence a une ES plus grande que chacune de ses composantes.

Le nombre total de degré de liberté (dl) vaut 84 + 85 − 2 soit 167. La valeur critique de t^* pour autant de degrés de liberté est 1,96. L'IC à 95 % de la différence est $78 \pm 30{,}0 \cdot 1{,}96$ et s'étend donc de 19 à 137 ml/kg. Nous pouvons être sûrs à 95 % que la réduction moyenne de la production de selles dans la population des nourrissons traités (en comparaison avec les nourrissons contrôles) se trouve dans cet intervalle.

[17] **Cette partie contient des équations nécessaires pour effectuer les calculs statistiques soi-même. Elle peut être passée sans perdre le fil de l'exposé.**

a) Que vaut l'IC à 95 % de la différence entre les niveaux moyens du facteur Y ?
b) Quelles hypothèses devez-vous faire pour répondre à cette question ?

2. Pullan et coll. ont étudié l'utilisation de nicotine transdermique pour traiter la colite ulcéreuse[20]. Le niveau de nicotine plasmatique de départ valait $0{,}5 \pm 1{,}1$ ng/ml (moyenne $\pm$ DS ; $n = 35$). Après 6 semaines de traitement, le niveau plasmatique valait $8{,}2 \pm 7{,}1$ ng/ml ($n = 30$).

   Calculez l'IC à 95 % de l'augmentation du niveau de nicotine plasmatique.

   Quelles hypothèses faites-vous ? Sont-elles respectées ?

   Si vous aviez accès à toutes les données, comment pourriez-vous les analyser ?

3. Au cours de plusieurs expériences, vous mesurez le nombre de récepteurs dans des cellules en culture en présence et en l'absence d'une hormone (chaque expérience a donc son propre contrôle). Les expériences ont été réalisées à plusieurs mois d'intervalle. Les cellules des différentes expériences avaient donc poussé dans des lots de sérums différents. Les résultats sont présentés dans le tableau suivant.

	Contrôle	Hormone
Expérience 1	123	209
Expérience 2	64	103
Expérience 3	189	343
Expérience 4	265	485

   Comment résumeriez-vous et représenteriez-vous graphiquement ces résultats ? Quel IC calculeriez-vous ?

4. Pourquoi l'ES d'une différence entre deux moyennes est plus grande que l'ES de chacune de ces moyennes ?

[20] RD Pullan, J Rhodes, S Ganesh, *et al.* Transdermal nicotine for ulcerative colitis. *N Engl J Med* 330:811-815, 1994.

3 LA DIFFÉRENCE ENTRE DEUX PROPORTIONS

Une façon de résumer les données de l'exemple 8.1 est de calculer la différence entre les deux proportions. La maladie a progressé chez 28 % des sujets traités par placebo et chez 16 % des sujets traités par AZT. Dans notre échantillon, la différence vaut 28% – 16% soit 12 %.

De façon plus générale, la différence entre deux proportions peut être calculée au moyen de l'équation 8.1 utilisant les variables A et D telles que définies dans le tableau 8.2. On appelle ce genre de tableau « tableau de contingence ».

$$\text{Différence entre deux proportions} = \Delta = p_1 - p_2 = \frac{A}{A+B} - \frac{C}{C+D} \qquad (8.1)$$

L'équation permettant de calculer l'IC à 95 % de la différence se trouve à la fin du chapitre. Dans cet exemple, l'IC à 95 % de la différence va de 6,7 % à 17,3 %. Si nous supposons que nos sujets sont représentatifs de la population des adultes infectés par le VIH encore asymptomatiques, nous sommes sûrs à 95 % que le traitement à l'AZT diminuera l'incidence de la progression de la maladie d'une valeur comprise entre 6,7 % et 17,3 %. Remarquons que ces calculs traitent de la vraie différence entre les incidences et pas du changement relatif. Lorsqu'on étudie un facteur de risque susceptible d'augmenter le risque de maladie, la différence entre les deux incidences est appelée le « risque attribuable »[23].

TABLEAU 8.2 Tableau de contingence

	Malades	Non malades	Total
Exposés ou traités	A	B	A + B
Non exposés ou placebo	C	D	C + D
Total	A + C	B + D	A + B + C + D

4 LE RISQUE RELATIF

Intuitivement, on pense plus souvent au rapport de deux proportions qu'à leur différence. Ce rapport est appelé le *risque relatif*. Le risque relatif est le rapport des incidences (équation 8.2)

$$\text{Risque relatif} = \frac{p_1}{p_2} = \frac{\frac{A}{A+B}}{\frac{C}{C+D}} \qquad (8.2)$$

[23] Le terme « risque attribuable » a été défini de quatre façons différentes. Lorsqu'on rencontre ce terme dans des publications, il faut s'assurer qu'on sait à quoi les auteurs font référence. Pour plus d'informations sur le risque attribuable, il faut lire le chapitre 4 de HA Kahn, CT Sempos. *Statistical Methods in Epidemiology*. New York, Oxford University Press, 1989.

8 LE RAPPORT DES COTES

Dans l'exemple de l'AZT, la cote de progression de la maladie chez les sujets traités à l'AZT vaut 0,19 et la cote de progression de la maladie chez les patients contrôles vaut 0,39. On peut résumer les données en prenant le rapport de ces valeurs, appelé le *rapport des cotes*. En comparaison avec les patients contrôles, la cote de progression de la maladie chez les sujets traités par l'AZT vaut 0,19/0,39 soit 0,49. En d'autres termes, la cote de progression de la maladie chez les patients traités à l'AZT vaut environ la moitié de celle des patients contrôles.

De façon plus générale, la cote de maladie chez les patients exposés (ou traités) se calcule comme suit (équation 8.4) :

$$\begin{aligned}\text{Cote de maladie chez les exposés} &= \frac{A/(A+B)}{B/(A+B)} = A/B \\ \text{Cote de maladie chez les non exposés} &= \frac{C/(C+D)}{D/(C+D)} = C/D\end{aligned} \tag{8.4}$$

$$\text{Rapport des cotes} = \frac{\text{Cote chez les exposés}}{\text{Cote chez les non exposés}} = \frac{A/B}{C/D} = \frac{AD}{BC} \tag{8.5}$$

Si une quelconque valeur de A à D vaut 0, on ne peut pas utiliser l'équation 8.5. Dans de telles situations, certains chercheurs ajoutent 0,5 à chacune des 4 valeurs A à D avant de calculer le rapport des cotes et son IC.

L'équation nécessaire pour calculer l'IC d'un rapport des cotes sera donnée plus loin dans le chapitre. Les calculs sont habituellement réalisés au moyen d'un ordinateur. L'IC du rapport des cotes de notre exemple s'étend approximativement de 0,36 à 0,67. L'interprétation ne devrait pas poser de problème. Si l'échantillon est représentatif de l'ensemble de la population, nous pouvons être sûrs à 95 % que le rapport des cotes de la population se trouve dans cet intervalle.

La plupart des personnes trouvent le concept de risque relatif facile à comprendre mais ils trouvent le rapport des cotes un peu déroutant. Lors de l'analyse d'études prospectives, transversales ou expérimentales, il n'y a pas d'avantage particulier à calculer un rapport des cotes. Comme nous le verrons dans le prochain chapitre, les rapports des cotes sont essentiels pour l'analyse des études rétrospectives cas-témoins.

9 LE RISQUE RELATIF DANS LES ÉTUDES DE SURVIE

Dans le chapitre 6, nous avons appris à représenter graphiquement une courbe de survie pour un groupe. Comment peut-on comparer la survie de deux groupes ? Calculer le risque relatif global est une façon de faire. Si le risque relatif vaut 2,1, cela signifie que le taux de décès des sujets d'un groupe vaut (en moyenne) 2,1 fois celui de l'autre groupe.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

6. Au moyen d'un livre et d'une calculatrice, vous devez être capables de calculer l'IC à 95 % d'un risque relatif ou d'une différence entre deux proportions.

EXERCICES

1. Le risque relatif de décès par cancer du poumon chez les fumeurs (en comparaison aux non-fumeurs) vaut environ 10. Le risque relatif de décès par maladies coronariennes chez les fumeurs vaut environ 1,7. Dans une population où la mortalité chez les non-fumeurs est de 5/100 000 par cancer du poumon et 170/100 000 par maladies coronariennes, le tabac cause-t-il plus de décès par cancer du poumon ou par maladies coronariennes ?
2. Goran-Larsson et coll. se sont demandé si l'hypermobilité des articulations causait des symptômes chez les musiciens[27]. Ils ont envoyé des questionnaires à de nombreux musiciens et ont posé des questions sur l'hypermobilité des articulations, les symptômes de douleur et de raideur. Ils ont posé des questions sur toutes les articulations mais cet exercice ne porte que sur les données qu'ils ont collectées au sujet des poignets. Parmi 96 musiciens avec des poignets hypermobiles, 5 % présentaient des douleurs et des raideurs des poignets. Par contre, 18 % des 564 musiciens sans hypermobilité présentaient de tels symptômes.
   a) Est-ce une étude prospective, rétrospective ou transversale ?
   b) Analysez les données de la façon la plus complète possible.
3. Un même nombre de cellules (100 000 par ml) a été placé dans quatre flacons. Deux lignées de cellules ont été utilisées. Certains flacons ont été traités avec des substances pharmaceutiques, tandis que les autres n'ont été traités qu'avec l'adjuvant (contrôle). Les données reprises dans le tableau suivant correspondent aux nombres moyens de cellules (milliers par millilitre) après 24 heures. Analysez ces données de la façon la plus complète possible. Si vous aviez accès à toutes les données originales, souhaiteriez-vous résumer les données autrement ? Comment ?

	Substance pharmaceutique	Pas de substance pharmaceutique
Lignée de cellules 1	145	198
Lignée de cellules 2	256	356

4. Cohen et ses collaborateurs ont étudié l'usage de la réanimation cardio-pulmonaire active (RCP)[28]. Dans la RCP standard, le réanimateur comprime le thorax de la victime pour forcer le cœur à pomper du sang vers le cerveau (et ailleurs) et ensuite relâche la pression pour permettre l'expansion du thorax. La RCP active se fait avec

[27] L Goran-Larsson, J Baum, GS Mudholkar, GD Lokkia. Benefits and disadvantages of joint hypermobility among musicians. *N engl J Med* 329:1079-1082, 1993.

[28] TJ Cohen, BG Goldner, PC Maccaro, AP Ardito, S Trazzera, MB Cohen, SR Dibs. A comparison of active compression-decompression cardiopulmonary resuscitation with standard cardiopulmonary resuscitation for cardiac arrests occuring in the hospital. *N Engl J Med* 329:1918-1921, 1993.

Si on introduisait ces nombres dans l'équation 8.2 pour calculer un risque relatif, la première étape consisterait à estimer le risque qu'une personne dont le chat a des puces contracte la maladie à $32/(32 + 4)$. Mais c'est un calcul inutile. Il se trouve que les auteurs ont choisi d'étudier un nombre égal de cas et de témoins. Ils auraient pu choisir d'étudier deux fois plus de témoins. Dans ce cas, ils auraient probablement trouvé 8 personnes se souvenant de puces parmi les 112. Le risque s'estimerait alors à $32/(32 + 8)$ et le risque relatif serait différent. Calculer le risque relatif dans des études cas-témoins conduit à un résultat insensé.

3 COMMENT INTERPRÉTER LE RAPPORT DES COTES ?

Comment pouvons-nous résumer les données d'une étude cas-témoins ? La réponse est étonnante. Alors que l'équation du risque relatif ne s'applique pas aux études cas-témoins, il est sensé de calculer le rapport des cotes comme dans les études prospectives. De plus, si la maladie est assez rare (touche seulement un faible pourcentage de la population étudiée), alors le rapport des cotes calculé dans une étude cas-témoins sera approximativement égal au vrai risque relatif. C'est une conclusion surprenante qui sera démontrée algébriquement plus loin. C'est une notion tellement importante qu'elle vaut la peine d'être répétée et soulignée : *si la maladie est assez rare, le rapport des cotes calculé dans une étude cas-témoins sera approximativement égal au vrai risque relatif.*

Dans cet exemple, le rapport des cotes vaut 17,3. Si nous supposons que la maladie des griffes du chat est rare parmi les propriétaires de chat[30], nous pouvons conclure que les propriétaires de chat porteurs de puces sont environ 17,3 fois plus susceptibles d'attraper la maladie des griffes du chat que les propriétaires de chat sans puce. Une équation permettant de calculer l'intervalle de confiance à 95 % sera présentée à la fin du chapitre. Dans cet exemple, l'IC à 95 % du rapport des cotes s'étend de 5,5 à 54,6. Dans la population, nous pouvons être sûrs à 95 % que le vrai rapport des cotes se trouve quelque part entre 5,5 et 54,6. Vu l'importante largeur de cet intervalle, certains pourraient dire que cette étude est peu concluante. D'autres pourraient faire remarquer que, même si le vrai rapport des cotes était proche de la borne inférieure de l'IC, il serait encore important (risque multiplié par cinq). Bien que l'étude soit trop petite pour déterminer le véritable rapport des cotes de façon très précise, elle montre clairement que le vrai rapport des cotes se trouve, avec une quasi certitude, loin de 1.

[30] Nous savons par d'autres données que c'est une hypothèse raisonnable. L'étude cas-témoins ne nous apprend rien sur le fait que la maladie des griffes du chat soit rare ou courante parmi les propriétaires de chat. Comme nous n'étudions que la population des propriétaires de chat, l'incidence de la maladie dans d'autres populations (ou la population générale) est hors propos.

uniformément, il se peut qu'ils aient, par inadvertance, davantage insisté auprès des cas.

- Les cas ont souffert de la maladie des griffes du chat. Ils étaient sans doute motivés à aider les chercheurs à en savoir plus sur la maladie. Il est simplement arrivé aux témoins de répondre au téléphone au mauvais moment. Ils n'avaient probablement jamais entendu parler de la maladie des griffes du chat et leur principale motivation était sans doute de terminer l'interview aussi vite que possible. Il se peut donc que les cas et les témoins ne donnent pas une information aussi précise et détaillée.
- Les patients ont été enrôlés dans l'étude uniquement par l'intermédiaire de leur médecin traitant. Les personnes qui ne développèrent qu'une forme légère de la maladie et qui n'ont pas recherché de soins médicaux n'ont pas été incluses dans l'étude. Toute personne malade ne cherche pas de soins médicaux. Cette étude sélectionne des personnes qui consultent un médecin lorsqu'elles sont légèrement malades. Ce critère de sélection n'a pas été appliqué aux témoins.
- Les témoins ont été enrôlés dans l'étude uniquement s'ils étaient à leur domicile lorsque les chercheurs ont composé au hasard leur numéro de téléphone. Cette méthode sélectionne des personnes qui restent beaucoup à la maison et des personnes qui ont plus d'une ligne téléphonique. Elle empêche la sélection des personnes qui n'ont pas le téléphone, qui voyagent beaucoup ou qui filtrent leurs appels avec des répondeurs automatiques. Aucun de ces critères de sélection n'a été appliqué aux cas.

Il serait facile de concevoir une étude pour contourner quelques-uns des problèmes énumérés ci-dessus mais on introduirait probablement de nouveaux problèmes dans le processus. Nous ne voulons pas non plus trop apparier les cas et les témoins car nous risquerions d'apparier pour la variable d'intérêt (si l'appariement portait sur le fait que les chats avaient des puces ou non, ils n'auraient pas pu se poser de questions sur l'association entre les puces et la maladie). Il est habituellement possible de trouver des explications alternatives pour interpréter les données d'études cas-témoins et ces explications sont souvent plausibles. Voici cinq explications alternatives des données de l'étude de la maladie des griffes du chat :

- La maladie des griffes du chat est causée par un organisme propagé par les puces. C'est l'hypothèse à laquelle croient les chercheurs.
- Les patients savent qu'ils ont attrapé la maladie de leur chat et ont donc une mémoire plus précise de leur chat. Puisqu'ils ont un intérêt personnel dans la maladie des griffes du chat, ils sont susceptibles de penser plus soigneusement aux réponses qu'ils donneront aux questions des enquêteurs. Les deux groupes de chats peuvent avoir exactement autant de puces mais les patients sont plus susceptibles de s'en souvenir que les témoins.
- Les puces sont plus courantes dans les régions rurales de l'état et les chats des régions rurales ont plus tendance à se faire mordre par des tiques. Peut-être que les tiques sont le véritable vecteur de la maladie.
- Les témoins ont plus tendance à rester à la maison (et étaient plus souvent chez eux quand les chercheurs ont appelé, cfr ci-dessus). Ils ont de meilleurs contacts

4. Vous devez être capables de distinguer, parmi les tableaux 2 × 2, ceux qui sont des tableaux de contingence de ceux qui ne le sont pas.
5. Sans consulter de livre, vous devez être capables de calculer un rapport des cotes.
6. Vous devez être capables de reconnaître des données pour lesquelles calculer le risque relatif n'a pas de sens.
7. Avec l'aide d'un livre et d'une machine à calculer, vous devez être capables de calculer l'IC à 95 % d'un rapport des cotes.
8. Vous devez être capables de reconnaître les données pour lesquelles une analyse pour données appariées est appropriée.

EXERCICES

1. Un rapport des cotes peut-il être plus grand que un ? Négatif ? Nul ?
2. Le logarithme du rapport des cotes peut-il être plus grand que un ? Négatif ? Nul ?
3. Gessner et ses collaborateurs ont étudié un excès de maladie dans une communauté d'Alaska[33]. Ils avaient l'impression qu'une des deux réserves d'eau de la ville distribuait trop de fluor, conduisant à l'intoxication au fluor. Ils ont comparé 38 cas avec 50 témoins. Trente-trois cas se rappelaient avoir bu de l'eau provenant du système de distribution n°1 alors que quatre témoins seulement avaient bu de l'eau provenant de ce système.

   Analysez ces données de la façon qui vous semble appropriée. Formulez les hypothèses à respecter.

   Comment ces chercheurs auraient-ils pu mener une étude prospective pour tester leur hypothèse ?

[33] BD Gessner, M Beller, JP Middaugh, GM Whitford. Acute fluoride poisoning from a public water system. *N Engl J Med* 330:95-99, 1994.

2 UN EXEMPLE SIMPLE : LA PRESSION ARTÉRIELLE D'ÉTUDIANTS EN MÉDECINE[34]

On souhaite tester l'hypothèse selon laquelle la pression artérielle systolique diffère entre les étudiants en médecine de première et de deuxième année (EM1 et EM2 respectivement). Le stress de la formation médicale augmente peut-être la pression artérielle.

Mesurer la pression artérielle de toute la promotion représenterait beaucoup de travail pour une étude préliminaire. Au lieu de cela, nous sélectionnons au hasard cinq étudiants de chaque promotion et mesurons leur pression artérielle systolique arrondie au 5 mmHg les plus proches.

$$\text{EM1} : 120, 80, 90, 110, 95$$
$$\text{EM2} : 105, 130, 145, 125, 115$$

Dans un premier temps, il faut regarder les données. Un graphique peut aider. La figure 10.1 représente les pressions artérielles de tous les individus. De toute évidence, la pression artérielle a tendance a être plus basse chez les étudiants de première année que chez ceux

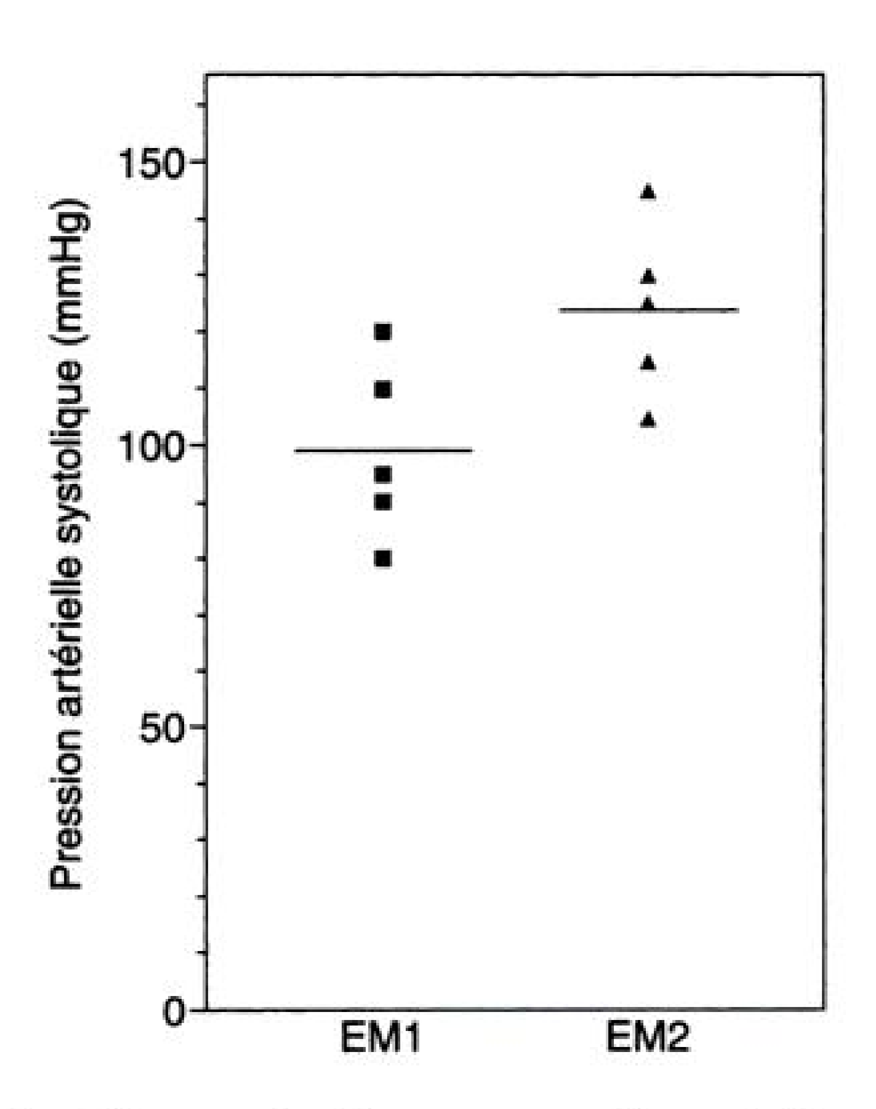


FIGURE 10.1 Données des échantillons représentées sur un graphique de dispersion en colonnes. Chaque carré représente la pression artérielle systolique d'un étudiant de première année (EM1). Chaque triangle représente la pression artérielle systolique d'un étudiant de deuxième année (EM2).

[34] Nous avons déjà rencontré ces données simulées dans l'exemple 5.2 et nous les retrouverons encore plus loin dans le livre.

- Lorsqu'on compare deux courbes de survie, l'hypothèse nulle consiste à dire que les deux courbes de survie sont les mêmes dans la population.

4 DE MAUVAISES INTERPRÉTATIONS FRÉQUENTES DE LA P-VALEUR

La P-valeur est facilement mal interprétée. Le meilleur moyen d'éviter cela est de garder à l'esprit ce qu'elle signifie : la P-valeur est la probabilité d'obtenir une différence aussi grande (ou plus grande) que celle observée si l'hypothèse nulle est vraiment correcte. Donc, une P-valeur de 0,03 signifie que, même si les deux populations avaient les mêmes moyennes, 3 % des études semblables à celle qui a été menée rapporteraient une différence au moins aussi grande que celle qui a été observée.

Il est très tentant d'extrapoler et de dire, « Oh, et bien, si il y a seulement 3 % de chance que la différence ait été causée par le hasard, alors, il doit y avoir 97 % de chance qu'elle résulte d'une véritable différence ! » C'est une erreur ! Ce qu'on peut dire c'est que si l'hypothèse nulle était vraie, 97 % des études aboutiraient à une différence plus petite que celle qui a été observée et 3 % des études aboutiraient à une différence aussi grande ou plus grande.

Le calcul de la P-valeur suppose que l'hypothèse nulle est correcte. La P-valeur ne peut pas dire si c'est correct. La P-valeur nous dit à quel point il serait rare d'observer une différence aussi grande ou plus grande que celle qui a été observée si l'hypothèse nulle était vraie. Le scientifique doit répondre à la question suivante : le résultat est-il tellement peu probable que l'hypothèse nulle peut être écartée ?

5 LES P-VALEURS UNILATÉRALES VERSUS BILATÉRALES

Une P-valeur bilatérale est la probabilité (supposant que l'hypothèse nulle est vraie) d'observer dans un échantillon aléatoire une différence au moins aussi grande que la différence observée, quel que soit le groupe ayant la plus grande moyenne. Une P-valeur unilatérale, par contre, est la probabilité (supposant que l'hypothèse nulle est vraie) d'observer dans un échantillon aléatoire une différence au moins aussi grande que la différence observée et que ce soit le groupe spécifié au préalable dans l'hypothèse de recherche qui ait la plus grande moyenne.

Si la différence observée va dans le sens prévu par l'hypothèse de recherche, la P-valeur unilatérale vaut la moitié de la P-valeur bilatérale[35]. Les termes P-valeur *unilatérale* ou *bilatérale* ont la même signification que les termes *une direction* ou *deux directions* parfois utilisés.

---

[35] Il existe des exceptions, comme le test exact de Fisher.

7 EXEMPLE 10.2. LA COMPARAISON DE DEUX PROPORTIONS DANS UNE ÉTUDE CAS-TÉMOINS

Il s'agit d'une étude cas-témoins investiguant l'association entre les puces et la maladie des griffes du chat. Les résultats ont été présentés dans le tableau 9.1.

Le rapport des cotes vaut 17,3. Nous voulons calculer une P-valeur. L'hypothèse nulle consiste à dire qu'il n'y a pas d'association entre les puces et la maladie des griffes du chat, que les chats des cas sont tout aussi susceptibles d'avoir des puces que les chats des témoins. La P-valeur bilatérale répond à cette question : si l'hypothèse nulle est vraie, quelle est la probabilité de choisir aléatoirement des sujets de sorte que le rapport des cotes soit égal ou plus grand que 17,3 ou égal ou plus petit que 0,058, l'inverse de 17,3 ?

Pour calculer une P-valeur, nous pourrions utiliser le test exact de Fisher ou le test du chi-carré. Les mêmes méthodes sont utilisées pour analyser des données d'études prospectives et d'études cas-témoins. Si un ordinateur réalise le travail, le test exact de Fisher sera le meilleur choix. Les résultats obtenus avec Instat sont présentés dans le tableau 10.2.

La P-valeur est de nouveau inférieure à 0,0001. S'il n'y a pas d'association entre les puces et la maladie des griffes du chat dans la population, il y a moins de 0,01 % de chance de tirer au hasard des sujets et de trouver une telle association.

Pour interpréter la P-valeur d'une étude cas-témoins, les hypothèses suivantes doivent être respectées :

- Les cas et les témoins sont sélectionnés aléatoirement dans leur population respective ou sont au moins représentatifs de ces populations.
- Chaque sujet a été sélectionné indépendamment des autres. Choisir un sujet ne devrait pas influencer la probabilité de choisir n'importe quel autre.
- À l'exception de l'absence de la maladie, les témoins ne diffèrent systématiquement des cas sur aucun point. Il est très difficile d'être sûr que cette hypothèse soit respectée.

Tableau 10.2 Résultats de l'exemple 10.2 obtenus avec Instat

Test exact de Fisher
La P-valeur bilatérale est < 0,0001, ce qui considéré comme très hautement significatif.
Il y a une association significative entre lignes et colonnes.
Rapport des cotes = 17,333
Intervalle de confiance à 95 % : 5,506 à 54,563 (en utilisant l'approximation de Woolf)

Comme on s'y attend, l'hypothèse nulle consiste à dire que les deux populations ont des courbes de survie identiques et que la différence observée dans nos échantillons est due au hasard. En d'autres termes, l'hypothèse nulle consiste à dire que le traitement par trois médicaments n'est pas meilleur (et n'est pas plus mauvais) que le traitement par deux médicaments pour prévenir les RGCH. La P-valeur du test du log-rank répond à la question suivante : si l'hypothèse nulle est vraie, quelle est la probabilité d'obtenir des courbes de survie aussi différentes avec des sujets sélectionnés aléatoirement ?

Les calculs sont assez compliqués et devraient être laissés aux logiciels informatiques. Les auteurs rapportent une P-valeur de 0,02. Si l'ajout d'un troisième médicament ne modifiait pas le développement de RGCH, on ne constaterait une telle différence entre les courbes de survie que dans 2 % des études de cette taille.

Nous avons déjà appris les hypothèses qui devaient être respectées pour interpréter une courbe de survie dans le chapitre 6. Le test du log-rank dépend des mêmes hypothèses (reprises ci-dessous) :

- Les sujets sont représentatifs de l'entièreté des bénéficiaires de greffe de moelle osseuse.
- Les sujets ont été choisis indépendamment.
- Critères invariables. Les critères d'inclusion et la définition de la *survie* doivent être constants tout au long de l'étude.
- La survie des sujets censurés devrait être la même, en moyenne, que la survie des autres sujets.

Les données de la figure 10.3 ne montrent qu'une seule réponse — le délai de survenue d'un RGCH. Les chercheurs ont aussi comparé d'autres variables, comme les fonctions rénale et hépatique, le temps de survie et la récidive de leucémie. L'analyse de toutes ces données suggère que l'ajout du troisième médicament aide à prévenir les RGCH sans causer d'autre problème. La recommandation d'une association thérapeutique de trois médicaments par les auteurs est basée sur l'analyse de toutes ces variables et pas seulement sur les courbes de survie.

RÉSUMÉ

La plupart de tests statistiques calculent une P-valeur. Même sans connaître les détails de tous les tests, il est essentiel que toute personne lisant la littérature biomédicale comprenne ce qu'est une P-valeur (et ce qu'elle n'est pas). Une P-valeur est simplement une probabilité qui répond à la question suivante : si l'hypothèse nulle est vraie (c'est-à-dire, si il n'y a pas de différence entre les populations), quelle est la probabilité qu'un échantillonnage aléatoire (compte tenu de la taille d'échantillon utilisée) aboutisse à une différence aussi grande ou plus grande que celle observée ?

5. Si la P-valeur est inférieure à α, conclure que la différence est *statistiquement significative* et décider de *rejeter l'hypothèse nulle*. Dans le cas contraire, conclure que la différence n'est pas *statistiquement significative* et décider de *ne pas rejeter l'hypothèse nulle*.

Remarquons que les statisticiens utilisent le terme *test d'hypothèse* d'une façon tout à fait différente des scientifiques. Tester des hypothèses scientifiques exige un travail assidu impliquant de nombreux types d'étude. Pour tester une nouvelle hypothèse, il est souvent nécessaire de concevoir une nouvelle méthodologie expérimentale et de concevoir des expériences contrôles ingénieuses. Il est par contre simple de tester une hypothèse statistique. Il suffit de vérifier si une P-valeur se trouve au-dessus ou en dessous d'un seuil.

Passer d'une P-valeur à la conclusion « significatif » ou « non significatif » rappelle les critiques de cinéma Siskel et Erbert. Dans leurs critiques écrites, ils évaluent chaque film sur une échelle (par exemple, trois étoiles et demi ou B−). C'est analogue à une P-valeur. C'est une manière concise de résumer leurs opinions sur le film. En critiquant des films à la télévision, ils prennent une décision pour le téléspectateur : « allez voir ce film ou n'y allez pas ».

La terminologie des tests d'hypothèse est plus facile à comprendre dans le cadre des contrôles de qualité. Par exemple, supposons que nous dirigeons une brasserie et que nous avons un entrepôt rempli de la dernière cuvée de bière. Avant de vendre cette cuvée, nous devons vérifier que la cuvée répond à certains standards de qualité. Plutôt que de tester toutes les bières de l'entrepôt (la population), nous choisissons aléatoirement quelques bouteilles (l'échantillon) pour les comparer à un « gold standard ». Les résultats peuvent se résumer par une P-valeur qui répond à la question suivante : si la nouvelle cuvée de bière est identique à la cuvée standard, quelle est la probabilité qu'un échantillon aléatoire de bouteilles soit aussi différent du standard que ce qui a été réellement observé ? Si la P-valeur est plus petite que α (habituellement 0,05), on rejette l'hypothèse nulle et donc la cuvée de bière (ou on réalise au moins d'autres tests pour trouver le problème). Si la P-valeur est plus grande que α, on ne rejette pas l'hypothèse nulle ni la cuvée.

2 LES AVANTAGES ET LES INCONVÉNIENTS DE L'EXPRESSION « STATISTIQUEMENT SIGNIFICATIF »

Utiliser l'expression *statistiquement significatif* présente trois avantages :

- Dans certaines situations, il est nécessaire d'arriver à une décision tranchée à partir d'une seule étude. On prendra une décision si les résultats sont significatifs et l'autre si les résultats sont non significatifs.
- Avec certains tests statistiques, il est difficile voire impossible d'obtenir une P-valeur exacte mais il est possible de déterminer si la P-valeur dépasse α ou non.
- Les gens n'aiment pas l'ambiguïté. La conclusion « les résultats sont statistiquement significatifs » est plus satisfaisante que la conclusion « un échantillonnage aléatoire produirait une différence au moins aussi grande dans 3 % des expériences si l'hypothèse nulle était vraie ».

- Nous dirigeons la phase III de l'essai clinique d'un médicament pour le traitement d'une maladie pour laquelle il n'existe pas de bon traitement. Si les résultats sont significatifs, nous commercialiserons le médicament. Si les résultats ne sont pas significatifs, les travaux sur ce médicament cesseront. Dans ce cas, une erreur de type I consistera à traiter de futurs patients avec un médicament inefficace, au lieu de ne pas les traiter. Une erreur de type II reviendra à suspendre le développement d'un bon médicament pour une maladie qu'on ne peut actuellement pas traiter. Ici, nous voulons fixer α à une valeur élevée parce qu'une erreur de type I ne porte pas à conséquences mais une erreur de type II serait inacceptable. On pourrait fixer α à 0,1.

Ce chapitre apporte volontairement peu de précisions sur les erreurs de type II et n'explique pas comment calculer cette probabilité. Nous apprendrons à le faire dans les chapitres 23 et 27.

Continuons notre analogie entre la signification statistique et le système juridique. L'équilibre entre erreur de type I et erreur de type II dépend du type de procès. Aux États-Unis (et dans beaucoup d'autres pays), un accusé dans une affaire criminelle est présumé innocent jusqu'à ce qu'il soit prouvé coupable « sans le moindre doute bien fondé ». Ce système repose sur la conviction qu'il est préférable de laisser de nombreuses personnes coupables en liberté que de condamner à tort une personne innocente. Le système est conçu de façon à éviter les erreurs de type I dans les affaires criminelles, même aux dépens de nombreuses erreurs de type II. On pourrait dire que α est fixé à une valeur très basse. Dans les affaires civiles, la cour ou le jury se prononce en faveur du plaignant si les preuves montrent que le plaignant « est plus susceptible d'être en droit que de ne pas l'être ». L'idée est que se prononcer à tort pour le plaignant n'est pas pire que de se prononcer à tort pour l'accusé. Le système cherche à égaliser les risques d'erreurs de type I et de type II dans les affaires civiles.

6 LA RELATION ENTRE α ET LES P-VALEURS

La P-valeur et α ont un rapport étroit. On calcule une P-valeur sur base des données. On fixe préalablement α sur base des conséquences des erreurs de type I et de type II. α est la P-valeur seuil sous laquelle une différence est appelée statistiquement significative.

7 LA RELATION ENTRE α ET LES INTERVALLES DE CONFIANCE

Bien que les intervalles de confiance (IC) et les P-valeurs soient présentées dans des parties différentes de ce livre, les deux sont étroitement liés. Ils sont tous deux basés sur les mêmes hypothèses et les mêmes principes statistiques. Il faut se demander si l'IC à 95 % contient la valeur énoncée dans l'hypothèse nulle. Si on compare deux moyennes, il faut se demander si l'IC à 95 % de la différence entre les moyennes contient 0. Si on analyse une

2 DES RÉSULTATS « HAUTEMENT SIGNIFICATIFS »

Intuitivement, nous pourrions penser qu'une P-valeur égale à 0,004 est plus significative qu'une P-valeur égale à 0,04. Au sens strict des définitions des termes, ce n'est pas correct. Une fois qu'une valeur a été fixée pour α, un résultat est soit significatif soit non significatif. Que la P-valeur soit très proche de α ou très éloignée n'a aucune importance. De nombreux statisticiens sont partisans de cette idée et pensent que le mot *significatif* ne devrait jamais être précédé d'un adjectif. La plupart des scientifiques sont moins stricts et mentionnent des résultats *très significatifs* ou *hautement significatifs* lorsque la P-valeur est minuscule.

Pour la représentation graphique de P-valeurs, les chercheurs utilisent communément une échelle de type « guide Michelin ». * $P < 0,05$ (significatif), ** $P < 0,01$ (hautement significatif), *** $P < 0,001$ (extrêmement significatif). Lorsqu'on lit ce genre de graphique, il faut connaître la légende des symboles puisque que la valeur seuil varie en fonction des chercheurs.

3 LES P-VALEURS LIMITES

Si nous suivons le modèle strict des tests d'hypothèse statistique et fixons α à sa valeur conventionnelle de 0,05, alors, une P-valeur de 0,049 indique une différence significative et une P-valeur de 0,051 indique une différence non significative. Cette distinction arbitraire est inévitable puisque tout l'intérêt de l'usage du terme *statistiquement significatifs* est d'arriver à une conclusion tranchée dans toute étude sans exception.

Plutôt que de regarder uniquement si le résultat est significatif ou non, il vaut mieux regarder la véritable P-valeur. De cette façon, nous saurons si la P-valeur est proche ou éloignée de α. Lorsqu'une P-valeur est juste un peu plus grande que α, certains scientifiques qualifient les résultats *de marginalement significatifs* ou *à la limite de la signification*.

Quand la P-valeur bilatérale est comprise entre 0,05 et 0,10, il est tentant de passer à une P-valeur unilatérale. La P-valeur unilatérale vaut la moitié de la P-valeur bilatérale et sera donc plus petite que 0,05. Les résultats deviennent significatifs comme par magie ! Évidemment, ce n'est pas une raison adéquate pour choisir une P-valeur unilatérale ! Il faut faire ce choix avant de collecter les données !

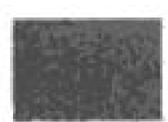
Un façon de s'en sortir avec une P-valeur limite serait de choisir entre trois décisions plutôt qu'entre deux. Plutôt que de décider si une différence est significative ou non significative, ajoutons une catégorie intermédiaire : *non concluante*. Cette approche n'est d'ordinaire pas utilisée.

4 LE TERME « NON SIGNIFICATIF »

Si la P-valeur est plus grande qu'une valeur α préalablement fixée, on dit que la différence est *non significative*. Cela signifie que les données ne sont pas assez puissantes

Les chiffres sont propres à cette étude mais le principe est universel. Toutes les études ont très peu de puissance pour détecter de minuscules différences et une puissance énorme pour détecter de grandes différences. Si on augmente le nombre de sujets, la puissance augmente.

Comme nous pouvons le constater, les calculs de puissance peuvent aider à interpréter des résultats non significatifs. Dans la littérature biomédicale, cependant, on rencontre rarement des calculs de puissance dans les articles qui présentent des résultats non significatifs. C'est en partie une affaire de tradition et en partie parce qu'il est difficile de définir la plus petite différence ou le plus petit risque relatif qu'on croit important.



RÉSUMÉ

Un résultat est statistiquement significatif quand la P-valeur est plus petite qu'une valeur préalablement fixée de α. Cela signifie que les résultats seraient surprenants si l'hypothèse nulle était vraie. L'usage statistique du mot *significatif* est assez différent de l'usage habituel de ce mot. Des résultats statistiquement significatifs peuvent être ou ne pas être scientifiquement ou cliniquement intéressants ou importants.

Des résultats statistiquement non significatifs signifient que les résultats ne sont pas incompatibles avec l'hypothèse nulle. Cela ne veut *pas* dire que l'hypothèse nulle est vraie. Lors de l'interprétation de résultats non significatifs, il peut être utile de regarder l'étendue de l'IC et de calculer la puissance de l'étude ou sa capacité à trouver un résultat significatif si les populations sont réellement différentes (avec une différence dont la taille est définie).

Tableau 13.1 Probabilité d'une petite P-valeur lorsqu'on teste de nombreuses hypothèses nulles

Nombre d'hypothèses nulles indépendantes (N)	2	3	4	5	6	7	8	9	10	20	50	100
Probabilité (P*) d'obtenir au moins une P-valeur inférieure à 0,05 par hasard	10 %	14 %	19 %	23 %	26 %	30 %	34 %	37 %	40 %	64 %	92 %	99 %
α* (Pour garder un α général à 0,05, il ne faut admettre qu'une P-valeur est significative que si elle est inférieure à cette valeur)	0,0253	0,0170	0,0127	0,0102	0,0085	0,0073	0,0064	0,0057	0,0051	0,0026	0,0010	0,0005

$P^* = 1 - 0{,}95^N$
$\alpha^* = 1 - 0{,}95^{(1/N)}$
Ce tableau suppose qu'α est fixé à sa valeur habituelle de 0,05. Pour calculer cette table pour d'autres valeurs, il suffit de remplacer « 0,95 » dans les deux équations par « $(1 - \alpha)$ ».

Si on ne dispose pas de ce tableau, il existe un moyen abrégé d'évaluer approximativement la valeur de la ligne du bas : il suffit de diviser 0,05 par le nombre de comparaisons. Dans cet exemple, le seuil est 0,05/3 soit 0,017 — la même valeur que celle du tableau, jusqu'à trois décimales. Si on testait 7 hypothèses, on calculerait par la méthode abrégée un seuil de 0,05/7 soit 0,0071 (qui est proche de la valeur exacte de 0,0073 du tableau). Si on fait plus de 10 comparaisons, cette méthode abrégée perd son utilité.

3 GROUPES MULTIPLES

3.1 Exemple 13.5

Hetland et ses collaborateurs se sont intéressés aux changements hormonaux des coureuses[42]. Parmi d'autres investigations, ils ont mesuré le niveau d'hormone luthéinisante (LH) chez des femmes non coureuses, des coureuses en amateur et des coureuses d'élite. Les niveaux hormonaux n'étant pas gaussiens, les chercheurs ont transformé leurs données en logarithmes de la concentration et ont effectué toutes les analyses sur les données transformées.

Même si cela paraît un peu douteux, c'est une bonne démarche à suivre puisque cela rend la population plus proche d'une distribution gaussienne. Les données sont présentées dans le tableau 13.2.

[42] ML Hetland, J Haarbo, C Christiansen, T Larsen. Running induces menstrual disturbances but bone mass is unaffected, except in amenorrheic women. *Am J Med* 95:53-60, 1993.

entre les deux courbes de survie était statistiquement significative avec une P-valeur inférieure à 0,025.

S'il s'agissait d'une étude réelle, il serait tentant de conclure que le traitement B est supérieur pour les patients les plus malades et de recommander le traitement B à ces patients dans l'avenir. Mais il ne s'agissait pas d'une véritable étude et les deux « traitements » reflétaient seulement une assignation aléatoire des patients. Les deux traitements étaient identiques; la différence observée était donc indubitablement liée au hasard. Il n'est pas surprenant que les auteurs aient trouvé une P-valeur basse parmi les six comparaisons faites. D'après le tableau 13.1, il y a 26 % de chance qu'une comparaison parmi six comparaisons indépendantes ait une P-valeur inférieure à 0,05 même si toutes les hypothèses nulles sont vraies. Pour réduire le risque général d'erreur de type I à 0,05, il faut réduire α à 0,0085 quand on compare six groupes.

C'est un problème difficile qui se présente souvent. Il faut se méfier des analyses de multiples sous-groupes car on risque fort de rencontrer des petites P-valeurs, même si toutes les hypothèses nulles sont vraies.

6 COMPARAISONS MULTIPLES ET DRAGAGE DES DONNÉES

Dans tous les exemples de ce chapitre, nous avons pu tenir compte de toutes les comparaisons multiples parce que nous connaissions toutes les comparaisons que les chercheurs avaient faites. Nous serions complètement trompés (et tirerions la mauvaise conclusion) si les chercheurs avaient fait de nombreuses comparaisons mais n'avaient publié que les quelques comparaisons significatives. Si l'hypothèse nulle est vraie, une P-valeur basse indique qu'une coïncidence rare est survenue. Mais on ne peut évaluer la rareté d'une coïncidence que si on sait combien de comparaisons différentes ont été faites. Comme nous l'avons vu, si on teste de nombreuses hypothèses nulles, la chance d'observer au moins une P-valeur significative est bien plus grande que 5 %. Par exemple, si on teste 100 hypothèses nulles indépendantes qui sont toutes vraies, nous avons 99 % de chance d'obtenir au moins une P-valeur significative. Nous serons complètement trompés si les chercheurs nous montrent la P-valeur significative et ne nous parlent pas des autres.

Pour éviter cette situation, les chercheurs devraient suivre ces règles :

- Les analyses doivent être planifiées avant la collecte des données.
- Tous les plans d'analyses devraient être établis consciencieusement, rapportés et respectés.

Les règles sont habituellement suivies religieusement dans les grands essais cliniques, particulièrement lorsque les données vont être réexaminées par la Food and Drug Administration. Mais ces règles sont souvent ignorées dans des études préliminaires plus informelles et dans la recherche en laboratoire. Dans de nombreuses situations, les chercheurs n'ont jamais réfléchi à la façon de réaliser les analyses avant de parcourir les données. Souvent, regarder les données suggère de nouvelles hypothèses à tester.

Il est difficile de savoir ce qu'il faut faire des analyses qui ne respectent pas ces règles. Si les chercheurs n'avaient pas décidé les hypothèses à tester avant de regarder leurs

14

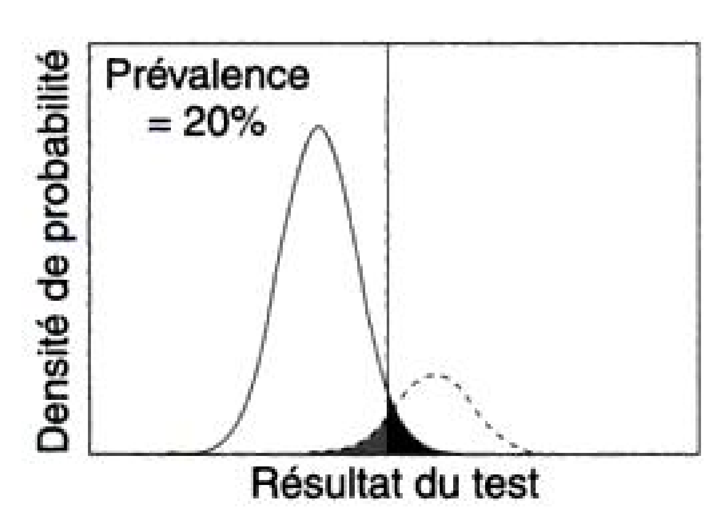
L'INTERPRÉTATION DE TESTS DIAGNOSTIQUES : INTRODUCTION À L'APPROCHE BAYÉSIENNE

Avis aux scientifiques en recherche fondamentale : il ne faut pas passer ce chapitre, même s'il paraît trop « clinique ». Ce chapitre monte le décor nécessaire à la discussion des deux chapitres suivants.

Quel est le rapport entre des tests diagnostiques et des P-valeurs ? Apprendre à interpréter des tests diagnostiques « positifs » et « négatifs » nous aidera à comprendre l'interprétation des tests statistiques « significatifs » et « non significatifs ».

1 LA PRÉCISION D'UN TEST DIAGNOSTIQUE QUALITATIF

Considérons tout d'abord un test qui apporte une réponse simple : positif ou négatif. Les résultats peuvent être présentés dans un tableau de contingence 2×2 (tableau 14.1). Les lignes représentent le résultat du test (positif ou négatif) et les colonnes indiquent la présence ou l'absence de la maladie (information obtenue sur base d'une autre méthode parfaitement exacte, sans doute le test meilleur test de l'époque). Si le test est « positif », il peut s'agir d'un test vraiment positif (vrai positif, VP) ou d'un test faussement positif (faux positif, FP) chez une personne qui n'a pas la maladie en question. Si le test est « négatif », il peut s'agir d'un test vraiment négatif (vrai négatif, VN) ou d'un test faussement négatif (faux négatif, FN) chez une personne atteinte de la maladie.



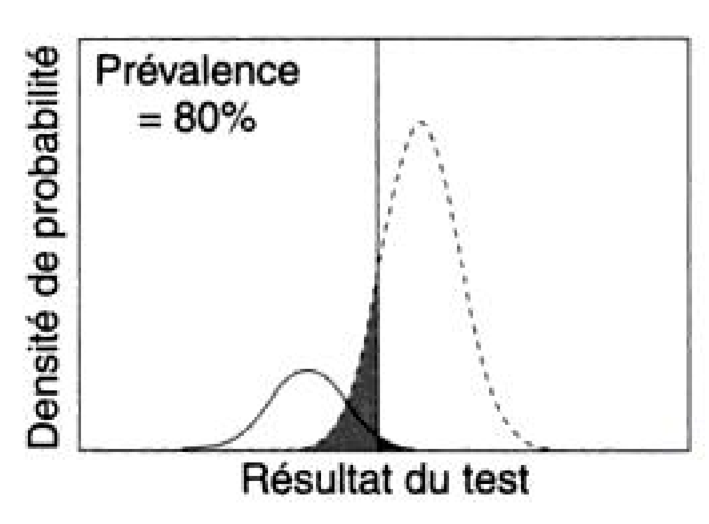


FIGURE 14.3 L'effet de la prévalence sur la valeur prédictive du test. Dans les figures 14.1 et 14.2, les deux courbes avaient la même surface, suggérant que la moitié de la population testée souffrait de la maladie. La partie gauche de cette figure montre les résultats obtenus dans le cas d'une prévalence de 20 % alors que la partie droite de l'image montre les résultats obtenus dans le cas d'une prévalence de 80 %. La proportion de tests positifs qui sont faussement positifs est plus importante dans la partie gauche. La proportion de tests négatifs qui sont faussement négatifs est plus importante dans la partie droite.

La sensibilité et la spécificité sont les caractéristiques du test. Les valeurs prédictives positive et négative, par contre, sont déterminées par les caractéristiques du test ainsi que par la prévalence de la maladie dans la population étudiée. Plus la prévalence est faible, plus le ratio des vrais positifs sur les faux positifs est bas.

Revenons à la figure 14.2. Les deux courbes ont la même surface suggérant qu'il y a autant de malades que de non malades parmi les personnes soumises au test. En d'autres termes, la figure 14.2 suppose que la prévalence est de 50 %. Dans la figure 14.3, la prévalence passe à 20 % (image de gauche) ou à 80 % (image de droite). Comme auparavant, la courbe continue représente les personnes saines et la courbe discontinue les malades. Tout résultat situé à droite de la ligne verticale pointillée est considéré comme positif. Il peut s'agir de vrais positifs (partie de la surface se situant en dessous de la courbe discontinue) ou de faux positifs (partie de la surface se situant en dessous de la courbe continue).

Les valeurs prédictives dépendent de la prévalence de la maladie. Une bien plus grande proportion de tests positifs sont des faux positifs dans la partie gauche de la figure que dans la partie droite. Par conséquent, la valeur prédictive d'un test positif est plus faible dans la partie gauche. Inversement, une bien plus petite proportion des tests négatifs sont des faux négatifs et par conséquent, la valeur prédictive d'un test négatif est bien plus grande dans la partie gauche.



4 LE CALCUL DE LA VALEUR PRÉDICTIVE D'UN TEST POSITIF OU NÉGATIF

La porphyrie aiguë intermittente est une maladie autosomique dominante difficile à diagnostiquer sur base de critères cliniques. Le diagnostic peut être posé en cas de diminution de l'activité de la porphobilinogène désaminase. Mais l'activité de cet enzyme

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

Tableau 14.6 Porphyrie, exemple D

	Maladie présente	Maladie absente	Total
< 79 unités	219	2	221
> 79 unités	281	498	779
Total	500	500	1000

valeur basse n'est pas normale, un plus faible niveau d'activité enzymatique est associé à une sensibilité plus faible et une spécificité supérieure. Peu de malades et très peu de sujets normaux présentent un niveau d'activité enzymatique aussi bas. Pour ce niveau, la sensibilité vaut 43,8 % et la spécificité 99,5 %. Si on examinait 1000 frères et sœurs de patients atteints de porphyrie, on s'attendrait à observer les résultats présentés dans le tableau 14.6.

On trouverait seulement 221 personnes avec un niveau enzymatique aussi bas parmi lesquelles 219 auraient la maladie. La valeur prédictive d'un test positif vaudrait 219/222 ou 98,6 %. Comme on pouvait s'y attendre, une activité enzymatique plus faible (un résultat de test plus anormal) a une valeur prédictive plus élevée.

5 LE THÉORÈME DE BAYES

Dans ces exemples, il a fallu parcourir plusieurs étapes pour calculer les valeurs prédictives à l'aide des tableaux. Ces étapes peuvent être intégrées dans une unique équation formulée par Thomas Bayes, un pasteur anglais qui a développé les mathématiques de la probabilité conditionnelle à la fin du XVIIIe siècle. L'équation peut s'écrire en termes de probabilité ou de cotes mais l'équation exprimée en termes de cotes est plus simple. Il faut donc se remémorer la différence entre les probabilités et les cotes avant d'apprendre le théorème de Bayes.

6 UNE RÉVISION DES PROBABILITÉS ET DES COTES

Une vraisemblance peut s'exprimer en termes de probabilité ou de cote.

- La **probabilité** qu'un événement se réalise est la proportion de fois où l'on s'attend à voir cet événement se réaliser au cours de plusieurs essais.
- La **cote** est définie comme la probabilité que l'événement se produise divisée par la probabilité que l'événement ne se produise pas.

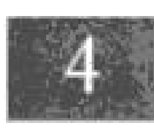
Une probabilité est une proportion et se situe toujours entre 0 et 1. Une cote se situe entre zéro et l'infini. N'importe quelle probabilité peut être exprimée en cote. N'importe quelle cote peut être exprimée en probabilité. Les cotes peuvent être converties en probabilités (et inversement) avec les équations 14.3 et 14.4 :

EXERCICES

1. Un test a une spécificité de 92 % et une sensibilité de 99 %. Calculez les valeurs prédictives d'un test positif et d'un test négatif dans une population où 5 % des individus sont malades.
2. Un test a une spécificité de 92 % et une sensibilité de 99 %. Calculez les valeurs prédictives d'un test positif et d'un test négatif dans une population où 0,1 % des individus sont malades.
3. Une femme veut savoir si son fils unique est daltonien. Son père étant daltonien, elle doit être porteuse du gène de cette maladie (le daltonisme est lié au sexe). Cela signifie qu'en moyenne, la moitié de ses fils seront daltoniens (elle n'a pas d'autre fils). Son fils est un petit enfant éveillé. Mais si vous lui demandez la couleur d'un objet, il semble répondre au hasard. Il ne maîtrise pas le concept de la couleur. Est-il daltonien ? Ou n'a-t-il pas encore compris la notion de couleur ? De votre expérience avec d'autres enfants du même âge, vous estimez que 75 % d'entre eux répondent correctement à des questions portant sur la couleur d'un objet. Combinez les antécédents génétiques de cet enfant et votre estimation sur les enfants de son âge pour calculer la probabilité qu'il soit daltonien.
4. Pour le patient C de l'exemple de la porphyrie, quelle serait la valeur prédictive d'un test positif si, par votre intuition clinique, vous estimiez la probabilité a priori à 75 % ?

réponse n'est pas nécessairement 5 %. Les statistiques conventionnelles sont incapables de répondre à cette question. La logique de Bayes peut y répondre mais seulement si on peut définir la probabilité a priori que l'hypothèse nulle soit vraie. La partie suivante explique comment appliquer la logique de Bayes aux P-valeurs.

Dans chaque cas (tests diagnostiques et tests statistiques) la logique de la première question part de la population vers l'échantillon et la logique de la seconde question part de l'échantillon vers la population. Pour analyser des données, la deuxième question nous intéresse davantage.



4 LA VALEUR PRÉDICTIVE DE RÉSULTATS SIGNIFICATIFS : BAYES ET P-VALEURS

On effectue un test statistique et on obtient un résultat significatif. Reprise de la dernière partie, voici la question à laquelle on souhaite répondre :

Dans quelle proportion de l'ensemble des études qui donnent des P-valeurs significatives l'hypothèse nulle est-elle vraie ? Autrement dit, si un résultat est statistiquement significatif, quelle est la probabilité que l'hypothèse nulle soit vraie ?

Voici un exemple imaginaire. Des chercheurs d'une industrie pharmaceutique étudient certains médicaments potentiellement efficaces pour traiter l'hypertension. Ils testent ces médicaments sur un groupe d'animaux. Ils ont décidé qu'une baisse moyenne de 10 mmHg de pression artérielle était intéressante et ils utilisent des échantillons assez grands pour avoir une puissance de 80 %, capacité de trouver une différence significative ($\alpha = 0{,}05$) si la vraie différence entre les moyennes des populations est de 10 mmHg. (Nous apprendrons à calculer la taille d'un échantillon au chapitre 22).

Ils testent un nouveau médicament et constatent une baisse significative de la moyenne de pression artérielle. Il y a deux possibilités. Soit le médicament abaisse réellement la pression artérielle, soit le médicament ne modifie pas du tout la pression artérielle et ces chercheurs ont observé par hasard ces résultats chez les animaux traités. Dans quelle mesure chacune de ces deux possibilités est-elle probable ?

Puisque α est fixé à 0,05, on sait que 5 % des études réalisées avec des médicaments inactifs montreront une baisse significative de pression artérielle. Mais ce n'est pas la question posée. On veut connaître la réponse à une autre question : dans quelle proportion des études où s'observe une diminution significative de pression artérielle le médicament est-il réellement efficace ? La réponse n'est pas nécessairement 5 %. Pour calculer la réponse, il faut utiliser la logique de Bayes et considérer la probabilité a priori. La réponse dépend de ce qu'on connaît du médicament avant de commencer l'étude, exprimé par la probabilité a priori que le médicament soit réellement efficace. Ce point est illustré dans les trois exemples suivants.

L'approche bayésienne nous permet de comprendre pourquoi il faut interpréter les P-valeurs dans le contexte de ce que nous connaissons ou croyons déjà et pourquoi il faut tenir compte de la plausibilité biologique pour interpréter des données. Quand la théorie change, il est préférable d'adapter notre perception de la probabilité a priori et de changer notre interprétation des données. Par conséquent, si on demande à différentes personnes d'analyser les mêmes données, elles peuvent, en toute honnêteté, tirer des conclusions différentes en suivant la démarche appropriée.

6 L'APPLICATION INFORMELLE DE LA LOGIQUE DE BAYES

Dans les articles de recherche biomédicale, on rencontre rarement (voire jamais) les calculs de Bayes pour interpréter des P-valeurs. Peu de scientifiques utilisent les calculs de Bayes pour aider à l'interprétation des P-valeurs. Par contre, de nombreux scientifiques utilisent la pensée bayésienne de façon plus informelle sans formuler la probabilité a priori de façon explicite et sans réaliser de calculs supplémentaires. Prenant trois études à titre d'exemples, la réflexion pourrait se présenter comme suit :

- La première étude a testé une hypothèse biologiquement plausible et appuyée par des données antérieures. La P-valeur vaut 0,04, ce qui est marginal. On a le choix de croire que les résultats sont dus à une coïncidence qui se produit 1 fois sur 25 sous l'hypothèse nulle ou bien de croire que l'hypothèse de recherche est vraie. Puisque l'hypothèse est très crédible, on préférera la deuxième alternative. L'hypothèse nulle est sans doute fausse.
- La seconde étude a testé une hypothèse incompréhensible du point de vue biologique et qui n'a jamais été soutenue par des données antérieures. La P-valeur vaut 0,04, légèrement inférieure au seuil habituel de 0,05. On a le choix de croire que les résultats sont dus à une coïncidence qui arrive 1 fois sur 25 sous l'hypothèse nulle ou de croire que l'hypothèse de recherche est vraie. Puisque l'hypothèse de recherche est invraisemblable, on croira plus facilement que les résultats sont dus à une coïncidence. L'hypothèse nulle est sans doute vraie.
- La troisième étude a testé une hypothèse invraisemblable du point de vue biologique et qui n'a pas été soutenue par des données antérieures. On serait étonné si cela s'avérait vrai. La P-valeur est incroyablement basse (0,000001). Après avoir analysé les détails de l'étude, on ne peut identifier aucun biais ni aucune imperfection. Il s'agit de scientifiques de bonne réputation et ils ont sans doute rapporté leurs données honnêtement. On a le choix de croire que les résultats sont dus à une coïncidence qui survient une fois sur un million sous l'hypothèse nulle ou de croire que l'hypothèse de recherche est vraie. Même si l'hypothèse de recherche semble insensée, les données poussent à la croire. L'hypothèse nulle est probablement fausse.

Il faut interpréter les données expérimentales dans le contexte de la théorie et des données antérieures. C'est pourquoi différentes personnes peuvent honnêtement tirer des conclusions différentes sur base des mêmes données.

16

LE THÉORÈME DE BAYES EN GÉNÉTIQUE

1 LE THÉORÈME DE BAYES ET LE CONSEIL GÉNÉTIQUE

Un conseil génétique a pour objectif de déterminer la probabilité qu'un individu ait un profil génétique particulier.

1.1 Exemple 16.1

Une femme veut connaître les chances qu'elle soit porteuse du gène de la dystrophie musculaire de Duchenne, une maladie récessive lié au chromosome X. Étant donné que son frère et son oncle maternel ont tous les deux la maladie, il est évident que le gène se transmet dans sa famille et qu'il ne s'agit pas d'une nouvelle mutation. D'après son histoire familiale, sa mère doit être porteuse du gène. Cette femme avait donc 50 % de chance d'hériter de ce gène à la naissance.

Le fait que cette femme ait deux fils indemnes de la maladie diminue les chances qu'elle soit porteuse du gène. La logique bayésienne permet de combiner ces évidences (deux fils en bonne santé) et l'histoire familiale (50 % de chance d'être porteuse). Nous réaliserons d'abord les calculs étape par étape au moyen d'un tableau et nous utiliserons ensuite l'équation de Bayes. Le tableau 16.1 montre les résultats qu'on s'attendrait à observer si on examinait de nombreuses femmes ayant la même histoire familiale et deux fils. Les calculs seront expliqués plus loin.

Pour construire le tableau, il faut suivre ces étapes :

1. Fixons le total général du tableau à 1000. Il s'agit d'un nombre arbitraire puisque seules les proportions nous intéressent.
2. Nous savons que la moitié des femmes sont porteuses. Inscrivons donc $1/2 \cdot 1000$ soit 500 comme total pour chaque colonne.

EXERCICES

1. Dans l'exemple 16.1, supposez que la femme avait 3 fils non malades. Quelle est la probabilité qu'elle soit porteuse de la maladie ?
2. Si le lod score est −3, quelle est la probabilité que le marqueur soit lié à la maladie ?
3. Il est possible de calculer une P-valeur à partir de données d'une étude de liaison. Expliquez dans un langage clair ce que cela veut dire.
4. Vous faites un test t et obtenez une P-valeur de 0,032. Vous avez examiné assez de sujets pour pouvoir garantir que l'expérience avait une puissance de 80 % pour détecter une différence déterminée entre les moyennes avec $P < 0{,}05$. Y a-t-il un sens à calculer un rapport de vraisemblance ? Si oui, calculez le rapport et expliquez ce qu'il signifie.

être induit en erreur. L'axe X ne commence pas à 0, ce qui donne l'impression que la variabilité est plus grande qu'elle ne l'est en réalité.

Au vu du graphique, il existe une relation évidente entre les deux variables. En général, les individus dont les muscles ont plus d'acides gras poly-insaturés C20-22 sont aussi plus sensibles à l'insuline. Les deux variables varient simultanément — les statisticiens disent qu'il y a une *covariation* importante ou une importante *corrélation*.

La direction et l'importance de la corrélation linéaire peut être quantifiée au moyen d'un coefficient de corrélation, abrégé r. Sa valeur peut aller de -1 à 1. Si le coefficient de corrélation vaut 0, alors les deux variables ne varient pas du tout simultanément. Si le coefficient de corrélation est positif, les deux variables ont tendance à augmenter ou à diminuer simultanément. Si le coefficient de corrélation est négatif, les deux variables sont inversement associées, c'est-à-dire que quand une variable a tendance à diminuer, l'autre a tendance à augmenter. Si le coefficient de corrélation vaut 1 ou -1, les deux variables varient parfaitement simultanément, autrement dit, graphiquement, les données forment une ligne droite.

Dans l'exemple, les deux variables augmentent simultanément, donc le coefficient de corrélation doit être positif. Comme les données sont un peu dispersées, le coefficient de corrélation doit donc être inférieur à 1,0. En fait, le coefficient de corrélation est égal à 0,77[50]. Comme toujours, nous souhaiterions inférer le coefficient de corrélation à l'entièreté de la population. Nous connaissons le coefficient de corrélation pour cet échantillon de 13 hommes. En utilisant une équation donnée en fin de chapitre, nous pouvons calculer que l'intervalle de confiance (IC) à 95 % du coefficient de corrélation s'étend de 0,38 à 0,93. Nous pouvons être certains à 95 % que le coefficient de corrélation de la population se situe dans cet intervalle. Même l'extrémité la plus basse de l'IC représente une forte corrélation. Nous pouvons donc être assez confiants sur le fait qu'il existe une forte corrélation dans la population.

Il est possible de calculer une P-valeur sur base de ces données. L'hypothèse nulle est qu'il n'y a pas de corrélation dans la population. La P-valeur bilatérale répond à la question suivante : si l'hypothèse nulle est vraie, quelle est la probabilité que, pour 13 sujets choisis au hasard, r soit supérieur à 0,77 ou inférieur à $-0,77$?

Dans cet exemple, la P-valeur vaut 0,0021. Si il n'y avait vraiment pas de relation entre la sensibilité à l'insuline et le pourcentage d'acide gras C20-22, il y aurait seulement 0,21 % de chance d'observer par hasard une corrélation aussi forte dans une étude de cette taille.

2 INTERPRÉTATION DU r

Pourquoi les deux variables sont-elles si bien corrélées ? Il y a quatre explications possibles :

[50] Toutes les valeurs ont été calculées à partir des données extraites du graphique de l'article. Comme c'est un peu imprécis, les calculs présentés dans l'article sont légèrement différents.

La plupart des méthodes non paramétriques se basent sur une idée simple. Il faut ranger les valeurs par ordre croissant et attribuer un rang à chaque valeur. Toutes les analyses qui suivront seront basées sur les rangs. En analysant les rangs et non les valeurs, on ne doit plus se soucier de la distribution de la population. Une méthode non paramétrique permettant de quantifier la corrélation est la *corrélation des rangs de Spearman*. La corrélation des rangs de Spearman se base sur les mêmes hypothèses que la corrélation ordinaire (de Pearson) évoquée plus haut si ce n'est que la corrélation des rangs ne requiert pas de distributions gaussiennes. Pour l'exemple de la sensibilité à l'insuline, $r_s = 0{,}74$ avec un IC à 95 % allant de 0,31 à 0,92.

7 NE PAS COMBINER DEUX POPULATIONS EN CORRÉLATION

Pour interpréter les résultats d'une corrélation ou d'une régression linéaire, il faut s'assurer que toutes les données proviennent d'une seule population. Si les données sont issues de deux populations, on peut facilement être induit en erreur. La figure 17.3 donne un exemple. Quand toutes les données sont analysées, le coefficient de corrélation vaut 0,72. Les variables X et Y semblent présenter une corrélation extrêmement forte et la P-valeur vaut 0,0009. Mais les données proviennent en fait de deux populations différentes, représentées par des cercles et des triangles. Dans chacune des populations, les variables X et Y sont très faiblement corrélées, avec $r = -0{,}16$ et $r = 0{,}05$, respectivement. Chaque P-valeur est supérieure à 0,50. La conclusion correcte d'après ces données est que les deux

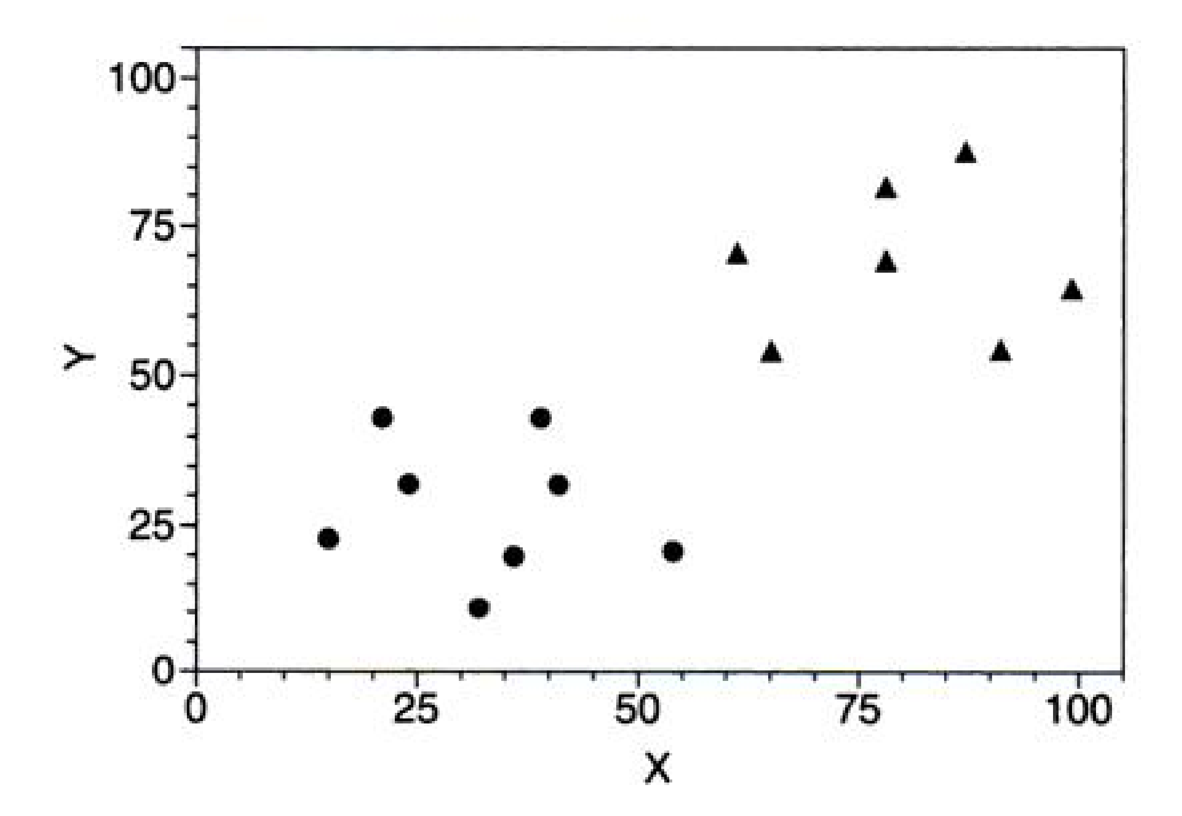


FIGURE 17.3 Il faut éviter de combiner deux populations dans une corrélation. Si on analyse toutes les données, on trouvera que $r = 0{,}72$ et $P = 0{,}0009$. Il semble très évident que X et Y sont corrélés. Mais on a en fait tiré l'échantillon dans deux populations différentes, représentées par des cercles et des triangles. Si on analysait chaque échantillon séparément, on trouverait des coefficients de corrélation très faibles et des P-valeurs supérieures à 0,50. Il n'y a en fait aucune évidence que X et Y sont corrélés. La combinaison des deux populations a créé l'illusion d'une corrélation.

3. Avec l'aide d'un livre, d'une machine à calculer ou d'un ordinateur, vous devez être capables de calculer le coefficient de corrélation entre deux variables. Vous devez également être capables d'obtenir la P-valeur appropriée.
4. Vous devez être capables de reconnaître les données pour lesquelles il est plus approprié d'utiliser le coefficient de corrélation de Spearman.
5. Vous devez être capables d'estimer le coefficient de corrélation d'un graphique XY.

EXERCICES

1. Dans l'exemple 17.1, comment les chercheurs auraient-ils analysé les données s'ils avaient mesuré deux fois la sensibilité à l'insuline et le pourcentage de C20-22 chez chaque sujet ?
2. La P-valeur de l'exemple 17.1 était bilatérale. Que vaut la P-valeur unilatérale ? Que signifie-t-elle ?
3. Faut-il mesurer les variables X et Y dans les mêmes unités pour calculer un coefficient de corrélation ? Peuvent-elles être mesurées dans les mêmes unités ?
4. Que vaut la P-valeur si r = 0,5 avec un échantillon de taille n = 10 ? Que vaut la P-valeur si r = 0,5 avec un échantillon de taille n = 100 ?
5. Pouvez-vous calculer un coefficient de corrélation si toutes les valeurs de X sont les mêmes ? Si toutes les valeurs de Y sont les mêmes ?
6. Golino et coll. ont étudié les effets de la sérotonine libérée au cours d'une angioplastie coronaire[53]. Après l'angioplastie (qui consiste à gonfler un ballon positionné dans la lumière d'une artère coronaire pour forcer l'ouverture d'une artère bouchée), ils ont mesuré le degré de vasoconstriction dans les premières minutes qui suivaient par le pourcentage de variation dans la surface de section (estimée au

Sérotonine (ng/ml)	% de changement de la surface de section
2,0	4,0
5,0	7,0
6,0	28,0
10,0	26,0
15,0	30,0
60,0	34,0
65,0	35,0
165,0	42,0

[53] P Golino, F Piscione, CR Benedict, *et al.* Local effect of serotonin released during coronary angioplasty. *N Engl J Med* 330:523-528, 1994.

courbe au graphique concentration versus réponse au dosage (qui peut être une densité optique, une radioactivité, une fluorescence, etc). Cette ligne ou courbe peut alors être utilisée pour déterminer la concentration à partir de la réponse obtenue avec des échantillons inconnus.

3 LES DIFFÉRENTS TYPES DE RÉGRESSION

La régression recouvre une grande famille de techniques.

3.1 Régression linéaire simple

C'est la forme la plus usuelle de régression. La variable dépendante est une mesure. Il y a une seule variable X. Le graphique de Y en fonction de X a l'allure d'une ligne droite. Le chapitre suivant discute de la régression linéaire.

3.2 Régression linéaire multiple

Y est toujours une variable quantitative (mesure) (pas une proportion et pas un temps de survie) mais il y a deux variables X ou plus. La régression multiple est appliquée pour déterminer l'influence d'une variable X tout en ajustant pour les effets des autres. La régression multiple (discutée dans le chapitre 31) est aussi utilisée pour trouver une équation qui permette de prédire des valeurs futures de la variable dépendante.

3.3 Régression logistique

Y est dans ce cas, une variable dichotomique (ou proportion) telle que, infecté/non infecté ou cancer/pas de cancer. Il peut y avoir une seule variable X, cependant la régression logistique est plus fréquemment utilisée avec plusieurs X. La régression logistique est discutée dans le chapitre 32.

3.4 Modèle de régression des risques instantanés proportionnels

La variable dépendante ici, est le temps de survie. Il peut y avoir une seule variable X, cependant le modèle de régression des risques instantanés proportionnels (discuté au chapitre 33) est plus fréquemment utilisé avec plusieurs X.

3.5 Régression non linéaire

À nouveau, Y est une mesure quantitative et il y a une seule variable X. Mais le graphe de Y en fonction de X est une courbe. La régression non linéaire est discutée dans le chapitre 34.

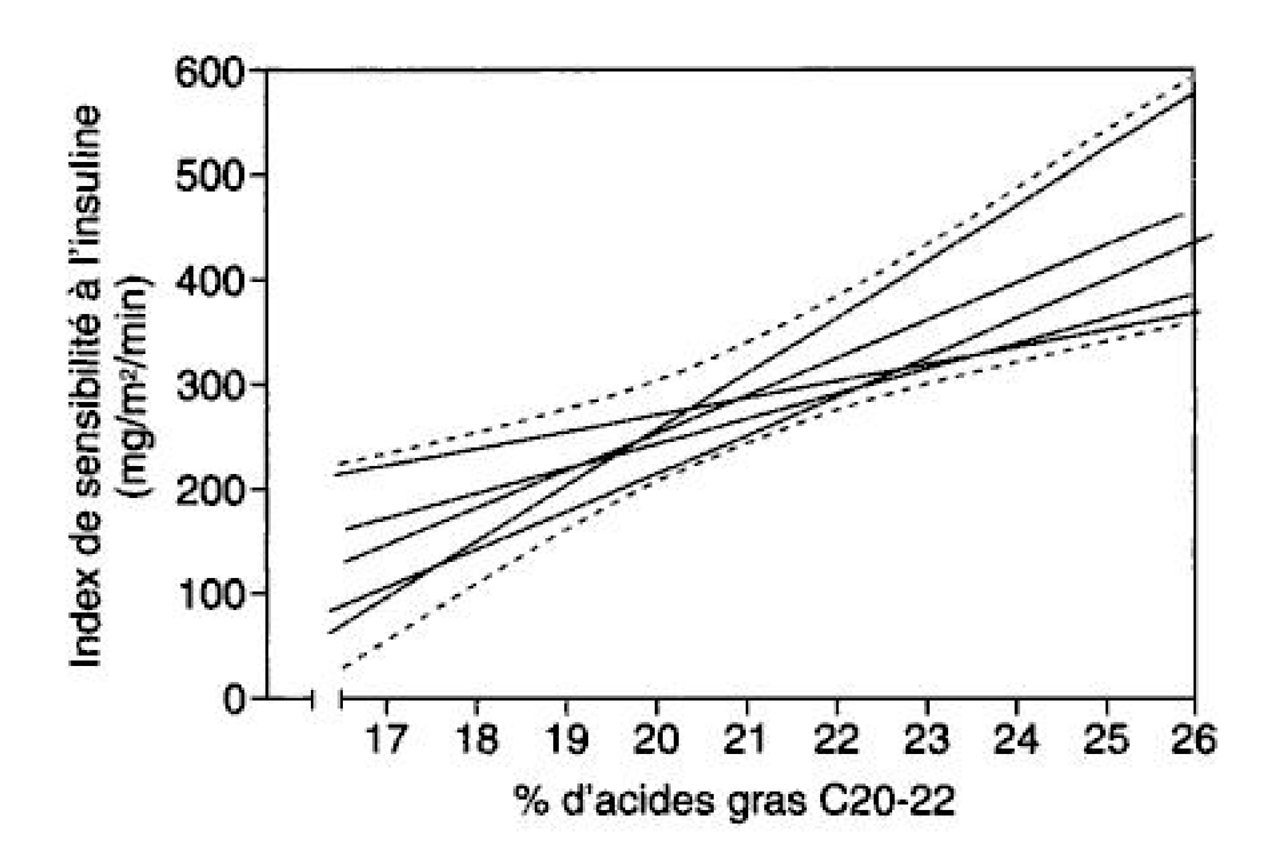


FIGURE 19.2 Interprétation de l'intervalle de confiance d'une droite de régression. Les limites de l'IC à 95 % de la droite de régression de notre exemple sont représentées par les courbes en pointillés. Si les conditions d'application de la régression linéaire sont acceptables, on peut être certain à 95 % que la droite réelle la mieux ajustée se trouve entre ces limites de confiance. Par exemple, cinq droites de régression possibles sont représentées (trait plein). Elles se trouvent toutes dans l'IC à 95 %.

Considérons à présent le tableau 19.1 qui nous présente les résultats numériques sortis par InStat.

- La pente vaut 37,2. Ceci signifie que lorsque le %C20-22 augmente de 1,0, la moyenne de la sensibilité à l'insuline augmente de 37,2 mg/m^2/min. Le programme fournit aussi l'IC à 95 % de la pente et son interprétation nous est familière. La pente calculée à partir d'un échantillon particulier n'est vraisemblablement pas égale à la pente réelle dans la population. Cependant, nous pouvons être certains à 95 % que la pente réelle se trouve dans le domaine défini par l'IC à 95 %, entre 16,7 et 57,7 mg/m^2/min.
- L'ordonnée à l'origine est −486,5. C'est la valeur de Y lorsque X vaut zéro. Pris au pied de la lettre, ceci signifie que s'il n'y avait pas d'acide gras C20-22 dans les membranes, la sensibilité à l'insuline serait −486,5. D'un point de vue biologique, ceci n'est pas possible puisque la sensibilité est la quantité de glucose nécessaire pour maintenir un taux sanguin constant et ne peut donc être négative. Nous discuterons ce problème plus loin.
- L'abscisse à l'origine est la valeur de X lorsque Y vaut zéro. Prise au pied de la lettre, cette valeur indique que lorsque le %C20-22 vaut 13,076, les muscles n'auront aucune sensibilité à l'insuline. Comme nous n'avons aucune donnée pour laquelle le %C20-22 est proche de 13, il n'y a pas moyen de savoir si c'est réellement le cas.
- Les résultats indiquent ensuite que r^2 vaut 0,5929. C'est la même valeur que celle déterminée par la corrélation linéaire au chapitre 17. Elle signifie que 59 % de la variabilité de la sensibilité en insuline peut être expliquée par le modèle de régression linéaire — que 59 % de la variabilité de la sensibilité à l'insuline peut être

5 RÉGRESSION LINÉAIRE : UNE MÉTHODE DE MOINDRES CARRÉS[56]

Comment la procédure de régression linéaire trouve-t-elle les « meilleures » valeurs de α et β pour faire en sorte que le modèle de régression (équation 19.2) s'ajuste aux données ? La régression linéaire définit la « meilleure droite » comme celle qui minimise la somme des carrés des distances verticales des points observés à la droite. Ceci implique que la somme serait plus élevée pour n'importe quelle autre droite.

Pourquoi minimiser la somme des carrés des distances ? La réponse la plus simple est que ceci permet de contourner le problème des distances négatives. Mais pourquoi ne pas minimiser la valeur absolue des distances ? Une réponse simple est que les distances sont élevées au carré parce qu'il est préférable d'avoir deux points relativement proches de la droite (disons 5 unités chacun) plutôt qu'un point très proche (1 unité) et le second très éloigné (9 unités). Une autre réponse est que la minimisation de la somme du carré des distances est un critère qui garantit que la droite dérivée de n'importe quel ensemble de données est unique. Une méthode qui minimise la valeur absolue des distances n'aurait pas toujours généré une réponse unique.

Il y a une autre manière de comprendre pourquoi la régression linéaire minimise le carré des distances. La distance verticale entre chaque point et la droite de régression est appelée *résidu*. La régression linéaire conduit à la droite qui minimise la DS des résidus. Lorsqu'on calcule la DS, on somme les carrés d'écarts ; donc pour minimiser la DS, il faut minimiser la somme des carrés d'écarts.

Remarquons que la régression linéaire ne conduit pas véritablement à la droite qui se trouve « la plus proche » des points car elle ne prend en considération que les distances verticales (parallèles à l'axe Y). Ceci implique également que les calculs de régression linéaire ne sont pas symétriques par rapport à X et Y. Une permutation de X et Y produira une droite de régression différente (à moins que les observations ne soient parfaites et que tous les points se trouvent exactement sur la droite). C'est compréhensible dans la mesure où l'objectif est de trouver la droite qui prédit le mieux Y à partir de X. La droite qui prédit le mieux X à partir de Y est le plus souvent différente.

Un exemple extrême permet d'éclaircir ce point. Considérons des données où X et Y ne sont pas du tout corrélés. On connaît X et on doit prédire Y. Le mieux que l'on ait à faire est de prédire que Y est égal à la moyenne des valeurs de Y pour toutes les valeurs de X. La droite de régression linéaire pour prédire Y en fonction de X est donc une droite horizontale à hauteur de la moyenne de Y. Inversement, la meilleure droite pour prédire X en fonction de Y serait une droite verticale située à la moyenne de X et à 90° de l'autre droite.

[56] Cette section est plus poussée que le reste. Elle peut être passée sans perte de continuité.

Quantifier l'exactitude de la valeur Y prédite est difficile. Pour réaliser cela, on souhaite calculer un intervalle de prédiction à 95 %, c'est-à-dire le domaine de valeurs qui contient 95 % des nouveaux points. Ceci est différent de l'IC présenté précédemment. L'intervalle de prédiction doit prendre en compte l'incertitude dans la position de la droite de régression (quantifiée par l'IC à 95 % de la droite de régression) et la dispersion des points autour de la droite. L'intervalle de prédiction est donc beaucoup plus large que l'IC.

La distinction entre IC et intervalle de prédiction est semblable à la différence entre erreur standard et DS et a déjà été discutée dans le chapitre 5. Les intervalles de prédiction sont toujours plus larges que les IC. Lorsque le nombre d'observations augmente, l'IC devient plus étroit tandis que l'intervalle de prédiction reste à peu près le même. Comme l'IC, l'intervalle de prédiction est curviligne.

10 LA RÉGRESSION VERS LA MOYENNE

Lorsqu'on interprète les résultats d'une régression linéaire, il faut s'assurer que les axes X et Y correspondent bien à des mesures distinctes. Sinon, on risque de faire fausse route à cause d'un problème nommé la *régression vers la moyenne*. Voici un exemple.

La figure 19.5 représente des données générées par ordinateur et simulant une expérience où la pression artérielle a été mesurée avant et après une intervention. Le diagramme de gauche présente les données. Chaque point est un individu dont la pression artérielle a été mesurée avant (axe X) et après (axe Y) une intervention. Les données sont

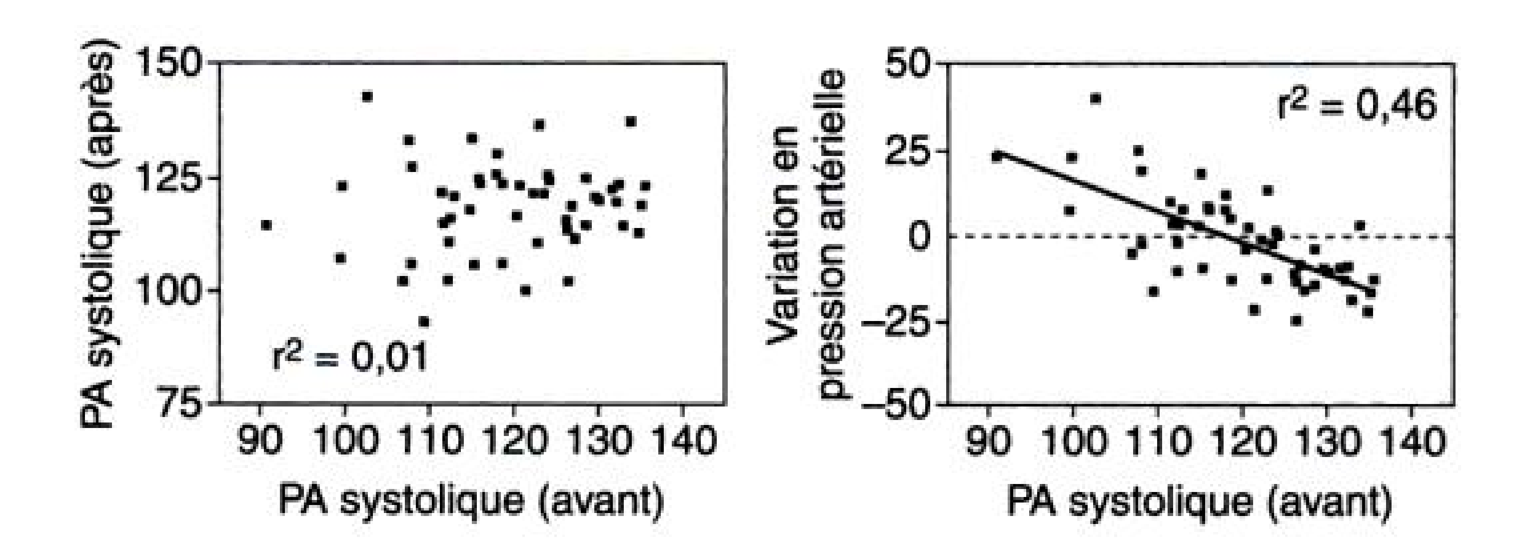


FIGURE 19.5 La régression vers la moyenne. Le graphique de gauche représente des données fictives de pression artérielle avant et après une intervention. Toutes les valeurs ont été échantillonnées d'une distribution gaussienne de moyenne 120 et de DS 10. Il n'y a pas de corrélation entre les deux ensembles de valeurs. Le graphique de droite présente les mêmes données après avoir effectué quelques manipulations. L'axe Y correspond à présent à la variation en pression artérielle (après — avant). Il apparaît une forte corrélation entre les deux variables et la droite de meilleur ajustement (représentée) a une pente qui est loin de l'horizontale. Les sujets qui au départ avaient une pression artérielle basse, avaient tendance à voir leur pression augmenter (partie gauche du graphique). Les sujets qui au départ avaient une pression artérielle élevée, avaient tendance à voir leur pression diminuer (partie droite du graphique). S'il s'agissait de données réelles, on pourrait être intrigué par ces résultats. Cependant, il n'y a aucun résultat à trouver et les données sont aléatoires. Parce que les valeurs portées en Y incluent les valeurs portées en X, la régression linéaire n'est pas appropriée pour analyser ces données.

RÉSUMÉ

La régression linéaire trouve la « meilleure » droite qui passe dans un nuage de points observés. Plus précisément, la régression linéaire trouve la droite qui minimise la somme des carrés des distances verticales des points à la droite.

OBJECTIFS

1. Vous devriez être familiers avec les termes suivants :
   - Régression
   - Modèle
   - Moindres carrés
   - Ordonnée à l'origine
   - Résidu
   - Pente
2. Vous devriez connaître les conditions d'application de la régression linéaire et savoir que l'utilisation de la régression linéaire peut conduire à des résultats erronés en cas de relation non linéaire entre les variables. Vous devriez savoir que les calculs de régression linéaire ne sont pas symétriques en X et Y.
3. En utilisant un livre, une machine à calculer et un ordinateur, vous devriez être capables de réaliser des régressions linéaires.
4. Vous devriez être capables d'interpréter les résultats de programmes de régression linéaire.
5. Vous devriez comprendre pourquoi les IC des droites de régression sont représentés par des courbes et pourquoi les intervalles de prédiction sont plus larges que les intervalles de confiance.

EXERCICES

1. La droite de régression sera-t-elle la même si vous permutez X et Y ? Qu'en sera-t-il du coefficient de corrélation ?
2. Pourquoi l'IC d'une droite de régression est-il curviligne ?
3. Les axes X et Y doivent-ils avoir les mêmes unités pour réaliser une régression linéaire ?
4. Combien de P-valeurs peut-on dériver d'une régression linéaire simple ?
5. Les résultats d'un dosage de protéines sont présentés dans le tableau ci-dessous. Des produits chimiques sont ajoutés à des tubes qui contiennent des quantités variables de protéines. La réaction provoque une couleur bleue. Les tubes avec une concentration plus élevée de protéines deviennent bleu foncé. La teinte plus ou moins foncée de la couleur bleue est mesurée par densité optique.

20

LE DESIGN DES ESSAIS CLINIQUES

Note aux scientifiques en sciences fondamentales : vous pouvez passer ce chapitre sans perdre le fil de l'exposé.

Il existe de nombreuses méthodes pour faire de la recherche médicale. Certaines recherches peuvent être faites en examinant des enregistrements existants tels que des certificats de décès ou des dossiers médicaux. D'autres types de recherche sont faits avec des animaux. Beaucoup d'études cliniques sont faites selon un design cas-témoins (voir chapitre 9). Ce chapitre traite des essais cliniques. Il s'agit d'expériences médicales où l'on donne un type de traitement à certains patients et un autre type à d'autres et les résultats sont comparés.

Bien que la médecine existe depuis des millénaires en tant que profession de la guérison, l'idée que les traitements médicaux doivent être testés de façon expérimentale est assez neuve. Quelques expériences médicales ont été publiées durant le XIXe siècle et au début du XXe siècle, mais ce n'est vraiment qu'après 1940 que la recherche clinique moderne s'est bien établie. Aujourd'hui il est généralement accepté que les nouveaux traitements soient testés avant d'être utilisés plus largement et il est aussi généralement accepté que le test soit prudemment contrôlé pour éviter des biais. Quand une anecdote ou une observation permet de formuler une hypothèse, on fait des expériences pour tester l'hypothèse. C'est une idée relativement récente en médecine. Cette idée devrait également s'imposer dans d'autres disciplines telle que l'éducation.

Avant qu'un nouveau traitement médicamenteux puisse être utilisé cliniquement, il doit être testé selon une série d'étapes bien définies. Au début, le travail est bien sûr pré-clinique. On peut déterminer de nombreuses propriétés du médicament en utilisant des animaux ou des cultures de cellules. Ensuite, la recherche clinique sur des nouveaux médicaments procède en quatre phases :

- On teste pour la première fois un médicament sur des êtres humains au cours de la phase 1. Le médicament est administré à quelques douzaines de personnes pour en

1.6 Le design d'inter-croisement

Dans le design d'inter-croisement, appelé aussi plan alterné ou plan avec permutation des traitements, chaque sujet est utilisé comme son propre contrôle. Ce design d'essai est utile pour des thérapies médicamenteuses qui peuvent être évaluées assez rapidement et pour des maladies qui ne progressent pas vite. Les designs d'inter-croisement ne peuvent habituellement pas être utilisés pour étudier des procédures chirurgicales (il n'est pas possible de « retirer » l'opération) ou pour une chimiothérapie cancéreuse (la mal progresse trop rapidement) mais sont utiles pour le traitement de conditions chroniques comme l'arthrite ou l'asthme. Pour ne pas introduire une nouvelle variable confondante (par exemple le temps ou l'ordre de traitement), l'ordre des deux traitements doit être appliqué aléatoirement pour chaque patient. Certains patients reçoivent d'abord le traitement A ; d'autres reçoivent d'abord le traitement B.

Deux problèmes se posent dans les designs d'inter-croisement. Premièrement, l'effet du premier traitement peut subsister durant la période du second traitement. Ces essais incluent donc généralement une période de sevrage entre les deux traitements pour éviter un effet résiduel. Le deuxième problème réside dans le fait que les sujets peuvent sortir de l'essai avant d'avoir reçu le second traitement.

1.7 Intention de traiter

Quand un sujet a été assigné aléatoirement à recevoir un certain traitement, il peut ne pas le suivre. Comment analyser les données de ce sujet ? À première vue, la réponse serait d'envisager le traitement que le patient reçoit réellement plutôt que le traitement qu'il est supposé recevoir mais cela pose des problèmes.

Supposons, par exemple, que nous comparons des traitements médicamenteux avec des traitements chirurgicaux (pontage coronaire) pour des patients souffrant d'une maladie cardiaque. Un des patients assigné à la chirurgie attrape une pneumonie et l'opération est donc annulée. Le patient ne reçoit pas le traitement chirurgical qui lui était assigné mais reçoit, au contraire, l'autre traitement. Admettons maintenant qu'il meure de pneumonie. Comment analyser ce décès ? On a spontanément envie de répondre que le patient a reçu un traitement médicamenteux et non chirurgical et donc que le décès doit être considérée comme un décès parmi les patients recevant le traitement médical. Mais cela biaiserait les résultats. La raison pour laquelle l'opération a été annulée était la pneumonie du patient. Si on retire ces patients du groupe chirurgical, alors on retire les patients les plus malades d'un groupe et on les additionne à l'autre. Les deux groupes ne sont plus comparables.

Une autre patiente a accepté de participer à l'essai sachant qu'elle peut être opérée ou qu'elle peut recevoir des médicaments. Elle est assignée à un traitement médicamenteux. Après réflexion, elle décide finalement de choisir la chirurgie. Comment analyser le résultat de cette patiente ? Elle n'a pas reçu le traitement qui lui était assigné. Ethiquement, on ne peut pas empêcher un patient de suivre le traitement qu'il souhaite. Scientifiquement, on ne veut pas considérer la réponse de cette patiente comme faisant partie du groupe chirurgical. Si chaque sujet choisit son propre traitement, alors les deux groupes vont différer sous de nombreux aspects et la comparaison des réponses sera dénuée de sens.

- Aucun sujet ne peut être enrôlé dans l'essai sans l'avoir formellement accepté en signant un document de consentement. Ce document explique le but de l'essai (en incluant une explication de la randomisation et du placebo si nécessaire) et énumère les avantages anticipés (potentiels) et les risques que prend le patient. Le document de consentement doit indiquer que la participation est volontaire et que le patient peut se rétracter sans pour autant perdre l'attention à laquelle il a droit. Ces documents doivent être écrits dans la langue du patient, avec des termes clairs et sans jargon médical. Habituellement, ils comptent deux à trois pages. Les comités d'éthique de la recherche portant sur des êtres humains ou du matériel humain permettent parfois aux chercheurs de court-circuiter l'information du consentement dans des circonstances inhabituelles (enfant en bas âge, patients comateux, traitements d'urgence) ou permettent de recourir au consentement des tuteurs ou des membres de la famille.
- Le sujet doit être tenu au courant d'éventuelles informations supplémentaires apparaissant au cours de l'essai et qui pourraient influencer sa décision de continuer l'essai.
- Les protocoles de recherche doivent être approuvés par un comité de cliniciens et de scientifiques ainsi que d'autres non impliqués dans la recherche. Ce procédé de révision permet d'éviter un enthousiasme excessif de chercheurs pour accomplir des essais dangereux. Dans la plupart des pays, ce procédé de révision est géré légalement. Le comité de révision est souvent appelé « Commission d'éthique ». Si la recherche porte sur un nouveau médicament, il doit d'abord être approuvé (aux USA) par la « Food and Drug Administration ». S'il porte sur des isotopes radioactifs ou des ADN-recombinants, il doit aussi être approuvé par des comités spécialisés.

5 L'ÉTHIQUE DU PLACEBO

On a parfois l'impression que les patients qui sont randomisés pour recevoir un placebo sont soumis à un mauvais traitement. Ils sont malades et, au lieu d'avoir un traitement expérimental, ils reçoivent un médicament inactif. Il n'est éthique pour les chercheurs d'utiliser un placebo dans un essai que si on pense qu'aucune des thérapies standards ne puisse être bénéfique au patient. Quand des thérapies classiques sont clairement efficaces, les patients du groupe contrôle doivent recevoir ces thérapies plutôt que le placebo.

Mais ce n'est pas parce qu'un traitement est standard, logique et accepté que cela veut dire qu'il aide les patients. Dans l'histoire médicale, il y a beaucoup d'exemples de thérapies qui étaient considérées à l'époque comme des standards et qui sont considérées comme nuisibles actuellement. George Washington a probablement été tué par saignée, une thérapie conventionnelle de son époque. L'oxygène pour les bébés prématurés est un autre exemple. Certains pensaient que l'oxygène pouvait être toxique et causer une forme de cécité (fibroplasie rétrocristallinienne). Pour tester cette idée, ils accomplirent plusieurs expériences contrôlées. À l'époque, on pensait qu'il n'était pas éthique de réduire l'oxygène (le traitement établi) délivré aux bébés prématurés atteints d'une maladie pulmonaire. Il

21

LES ESSAIS CLINIQUES AVEC n = 1

La plupart des essais cliniques engagent beaucoup de patients et le nombre de patients est abrégé « n ». Un essai avec n = 40 enrôle 40 patients. On peut répondre à certaines questions cliniques en faisant un essai formel avec un seul patient, n = 1.

L'objectif d'un essai où n = 1 est modeste. On n'essaie pas d'élucider les secrets de la nature et on n'essaie pas de rassembler des données qui aideront de futurs patients. On veut seulement savoir quelle est, parmi plusieurs alternatives, la meilleure pour le patient. Le médicament administré est-il meilleur qu'un placebo ? La dose A est-elle meilleure que la dose B ? Un médicament générique est-t-il plus efficace qu'un médicament de marque ?

Il n'est fondé de conduire un essai n = 1 que quand il existe un doute important au sujet de la meilleure thérapie à proposer. Peut-être que ni le patient, ni le clinicien ne pensent que la thérapie utilisée a de l'effet. On peut faire un essai n = 1 en comparant le médicament et le placebo pour éclairer la situation. Le patient insiste peut-être pour prendre un traitement que le clinicien croit sans valeur. On peut faire un essai n = 1 pour savoir si le médicament est meilleur que l'alternative. Un essai n = 1 est seulement faisable dans les conditions suivantes :

- La maladie est chronique et donc le traitement sera administré sur une longue période.
- L'état du patient est assez stable et donc la maladie ne s'aggravera pas au cours de l'essai.
- Les traitements font rapidement de l'effet et les effets sont rapidement réversibles.
- Le patient comprend le fondement de l'étude et veut y participer.
- Il est possible de quantifier l'efficacité du traitement expérimental par des mesures de laboratoire ou sur une échelle clinique (i.e. la qualité de vie). Idéalement, on enregistrera plusieurs variables qui mesurent l'efficacité du traitement ainsi que les effets secondaires.
- Un pharmacien est disponible pour préparer des médications non étiquetées et pour garder la trace du médicament et du placebo.

TABLEAU 22.1 Valeur de l'indice de puissance

α		Indice de puissance $(z_\alpha + z_\beta)^2$				
Unilatéral	Bilatéral	$\beta = 0,01$ Puissance = 99 %	$\beta = 0,05$ Puissance = 95 %	$\beta = 0,10$ Puissance = 90 %	$\beta = 0,20$ Puissance = 80 %	$\beta = 0,50$ Puissance = 50 %
0,05	0,10	15,8	10,9	8,6	6,2	2,7
0,025	0,05	18,3	13,0	10,5	7,9	3,8
0,005	0,01	23,9	17,8	14,9	11,7	6,6

Avant d'essayer de calculer l'indice de puissance pour différentes valeurs de z_α ou z_β, essayons d'abord de reproduire le tableau 22.1. Selon le tableau utilisé, il faudra considérer $z_1 - \alpha$, $z_1 - \beta$ ou $z_1 - \alpha/2$.

Exemple 22.3

Nous savons que la déviation standard de la pression artérielle dans notre population est d'environ 10 mmHg. De combien de sujets avons-nous besoin pour avoir une puissance de 80 % de détecter une différence de 5 mmHg entre les moyennes avec $P < 0,05$ (bilatéral) ? La réponse est environ 63 sujets dans chaque groupe.

Les programmes informatiques calculent parfois la taille d'échantillon pour plusieurs valeurs d'α et β. Le tableau 22.2 présente les résultats obtenus avec Instat pour l'exemple précédent.

TABLEAU 22.2 Calcul de la taille de l'échantillon pour l'exemple 22.3

Valeurs introduites Différence significative minimum que l'on souhaite détecter = 5 Déviation standard estimée pour chaque population = 10					
Résultats					
Puissance	β	α (bilatéral)			
		0.10	0.05	0.02	0.01
80 %	0.20	50	63	81	94
90 %	0.10	69	85	105	120
95 %	0.05	87	105	127	143

Les valeurs sont les nombres de sujets nécessaires dans chaque groupe.

RÉSUMÉ

Avant de commencer une étude, les chercheurs doivent connaître le nombre de sujets à y inclure. Il faut suffisamment de patients pour avoir des résultats valables mais il faut également éviter de perdre du temps et de l'argent en utilisant trop de sujets. On peut utiliser deux approches pour calculer la taille d'un échantillon. Une des approches consiste à déterminer le nombre de sujets nécessaires pour que l'intervalle de confiance (ou l'intervalle de confiance d'une différence) ait une largeur définie. L'autre approche consiste à déterminer le nombre de sujets nécessaires pour avoir une certaine puissance pour obtenir une différence significative (avec une valeur α spécifiée), étant donné une hypothèse de recherche spécifiée. Pour utiliser l'une ou l'autre méthode, il faut estimer la déviation standard des valeurs (si la réponse est une mesure) ou estimer la valeur de la proportion (si la réponse est binomiale).

OBJECTIFS

1. Sans utiliser de livre ni d'ordinateur, vous devez savoir quelles informations sont nécessaires pour calculer la taille d'un échantillon dans diverses situations.
2. En utilisant des livres, des machines à calculer ou des ordinateurs, vous devez être capables de déterminer la taille d'échantillon nécessaire pour obtenir une moyenne ou une proportion avec une certaine précision. Vous devez aussi être capables de calculer la taille d'échantillon nécessaire pour trouver une différence spécifiée avec une puissance spécifiée.

EXERCICES

1. Vous préparez le protocole d'une étude afin de tester si un nouveau traitement médicamenteux diminue la pression artérielle de façon importante. Par votre expérience, vous savez qu'il suffit de 15 rats dans chaque groupe. Préparez une analyse de puissance convaincante pour justifier cette taille d'échantillon.
2. L'incidence générale d'une maladie est de 1 pour 10 000. Vous pensez qu'un facteur d'exposition augmente le risque. De combien de sujets avez-vous besoin pour détecter un risque relatif aussi petit que 1,1 dans une étude prospective, avec 95 % de puissance ?
3. Quelle taille d'échantillon auriez-vous besoin pour réaliser un sondage électoral visant à déterminer, avec une erreur de 1 %, le pourcentage des voix destinées à chaque candidat ? Quelles hypothèses faites-vous ?
4. Une étude peut-elle atteindre une puissance de 100 % ?
5. Environ 10 % des patients meurent au cours d'une certaine opération. Une nouvelle technique pourrait réduire le taux de mortalité. Vous planifiez un essai pour

- Comme t est un rapport de deux grandeurs ayant les mêmes unités, il n'a pas d'unité.
- Il n'est pas nécessaire d'avoir les données brutes pour calculer t. Nous pouvons le calculer à partir de la moyenne, de l'erreur standard de la moyenne (ou de la déviation standard) et de la taille de chaque groupe.

2 COMMENT OBTENIR LA P-VALEUR À PARTIR DE t À L'AIDE D'UNE TABLE

Si l'on n'utilise pas un programme informatique qui calcule la P-valeur exacte, il faut déterminer la P-valeur à l'aide d'une table.

Pour convertir t en une P-valeur, nous devons pouvoir répondre aux trois questions suivantes :

- Quelle est la valeur de t ?

  Si elle est négative, on ne considère pas son signe (c'est-à-dire que l'on prend sa valeur absolue). Lorsque t augmente, la P-valeur diminue.
- Souhaite-t-on calculer des P-valeurs uni ou bilatérales ?

  Pour des raisons déjà mentionnées plus haut, ce livre utilise toujours des P-valeurs bilatérales.
- Combien de degrés de liberté (dl) y a-t-il ?

  Pour un test t appliqué dans le cas de deux échantillons, le nombre de dl est le nombre total de sujets moins deux. Pour l'exemple, dl = 8. Si l'on a pas étudié les statistiques de manière approfondie, les règles pour calculer les degrés de liberté n'étant pas toujours intuitives, il faut soit les apprendre, soit les rechercher. Pour toute valeur particulière de t, la P-valeur diminue lorsque le dl augmente.

TABLEAU 23.2 Détermination de la P-valeur à partir de t

dl	α (bilatéral)						
	0,250	0,200	0,150	0,100	0,050	0,010	0,001
6	1,273	1,440	1,650	1,943	2,447	3,707	5,959
7	1,254	1,415	1,617	1,895	2,365	3,499	5,408
8	1,240	1,397	1,592	1,860	2,306	3,355	5,041
9	1,230	1,383	1,574	1,833	2,262	3,250	4,781
10	1,221	1,372	1,559	1,812	2,228	3,169	5,587
15	1,197	1,341	1,517	1,753	2,131	2,947	4,073
20	1,185	1,325	1,497	1,725	2,086	2,845	3,850
100	1,157	1,290	1,451	1,660	1,984	2,626	3,390

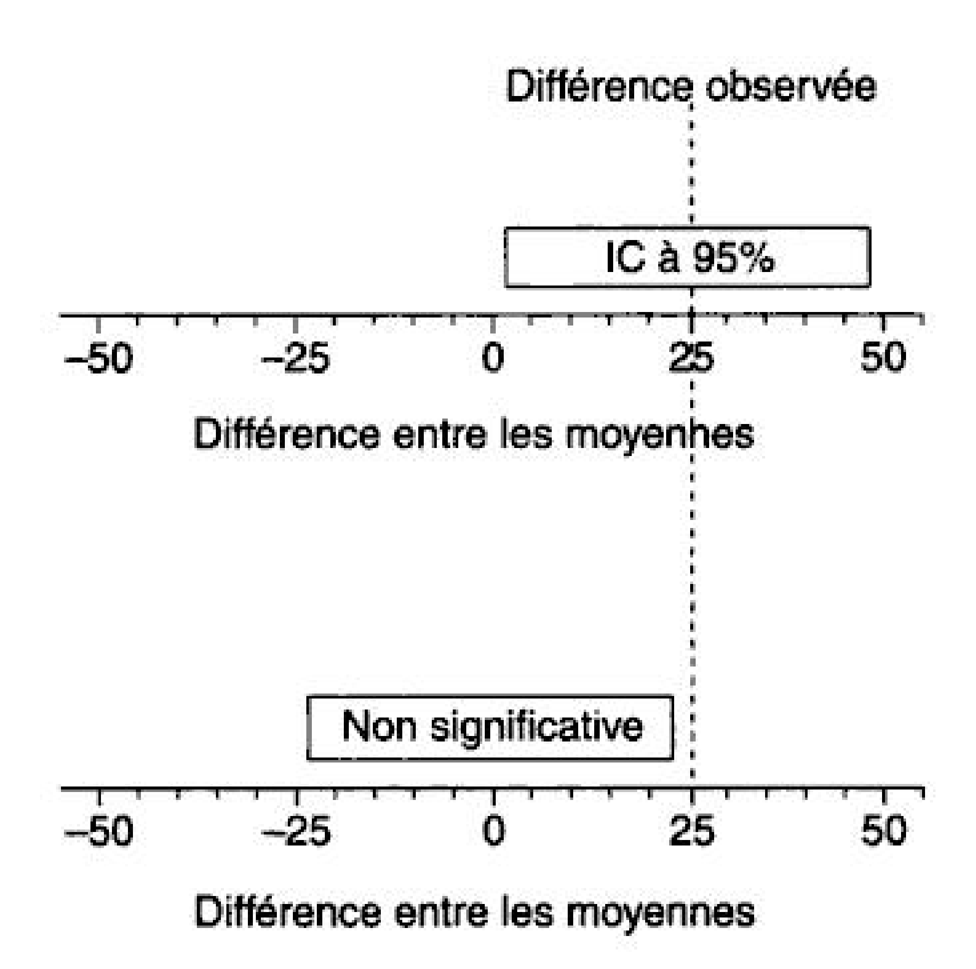


FIGURE 23.3 Relation entre IC et signification statistique. La largeur de l'IC à 95 % est égale à l'étendue du domaine où les différences ne seraient pas significatives ($\alpha = 0{,}05$). Dans cet exemple, l'IC à 95 % n'inclut pas 0 et le domaine des valeurs non significatives n'inclut pas la différence observée.

chaque direction sur une demi-largeur définie par l'équation 23.4 :

$$\text{Demi-largeur} = t^* \cdot \text{ES de la différence} \qquad (23.4)$$

Dans l'exemple traité, la valeur critique de t est 2,306 et l'ES de la différence est 9,84 mmHg. L'IC à 95 % s'étend sur une distance égale à 22,69 de chaque côté de la différence moyenne observée, égale à 25. Donc l'IC à 95 % s'étend de 2,31 à 47,69. Avec un degré de confiance de 95 %, on peut dire que la différence vraie entre les moyennes des populations se trouve dans cet intervalle représenté sous forme de boîte dans la partie supérieure de la figure.

Le résultat du test d'hypothèse se trouve quant à lui représenté dans la moitié inférieure de la figure. Étant donné la taille de l'échantillon et la dispersion, n'importe quelle différence entre les moyennes à l'intérieur de la boîte serait non significative et n'importe quelle différence à l'extérieur serait statistiquement significative. Déterminer où dessiner la droite est simple. On peut transformer l'équation 23.1 en l'équation 23.5 :

$$\text{Différence entre les moyennes} = t \cdot \text{ES de la différence} \qquad (23.5)$$

En se basant sur le tableau 23.2, on peut savoir que n'importe quelle valeur de t entre −2,306 et 2,306, conduira à une P-valeur (bilatérale) plus grande que 0,05 (si le nombre de degrés de liberté vaut 8). L'ES de la différence estimée à partir de nos données vaut 9,84. Donc une différence entre les moyennes des échantillons de $2{,}306 \times 9{,}84 = 22{,}69$ (dans n'importe quelle direction) sera exactement à la limite de la signification. La boîte correspondant à la « non-signification » s'étend d'une différence

TABLEAU 24.1 Réarrangements pour lesquels les valeurs moyennes du groupe A sont inférieures de 25 unités ou plus à celles du groupe B

Groupes	Valeurs	Différence entre les moyennes
A	80, 90, 95, 110, 120	25 (données observées)
B	105, 115, 125, 130, 145	
A	80, 90, 95, 105, 120	27
B	110, 115, 125, 130, 145	
A	80, 90, 95, 105, 110	31
B	115, 120, 125, 130, 145	
A	80, 90, 95, 110, 115	27
B	105, 120, 125, 130, 145	
A	80, 90, 95, 105, 115	29
B	110, 120, 125, 130, 145	
A	80, 90, 95, 105, 125	25
B	110, 115, 120, 130, 145	

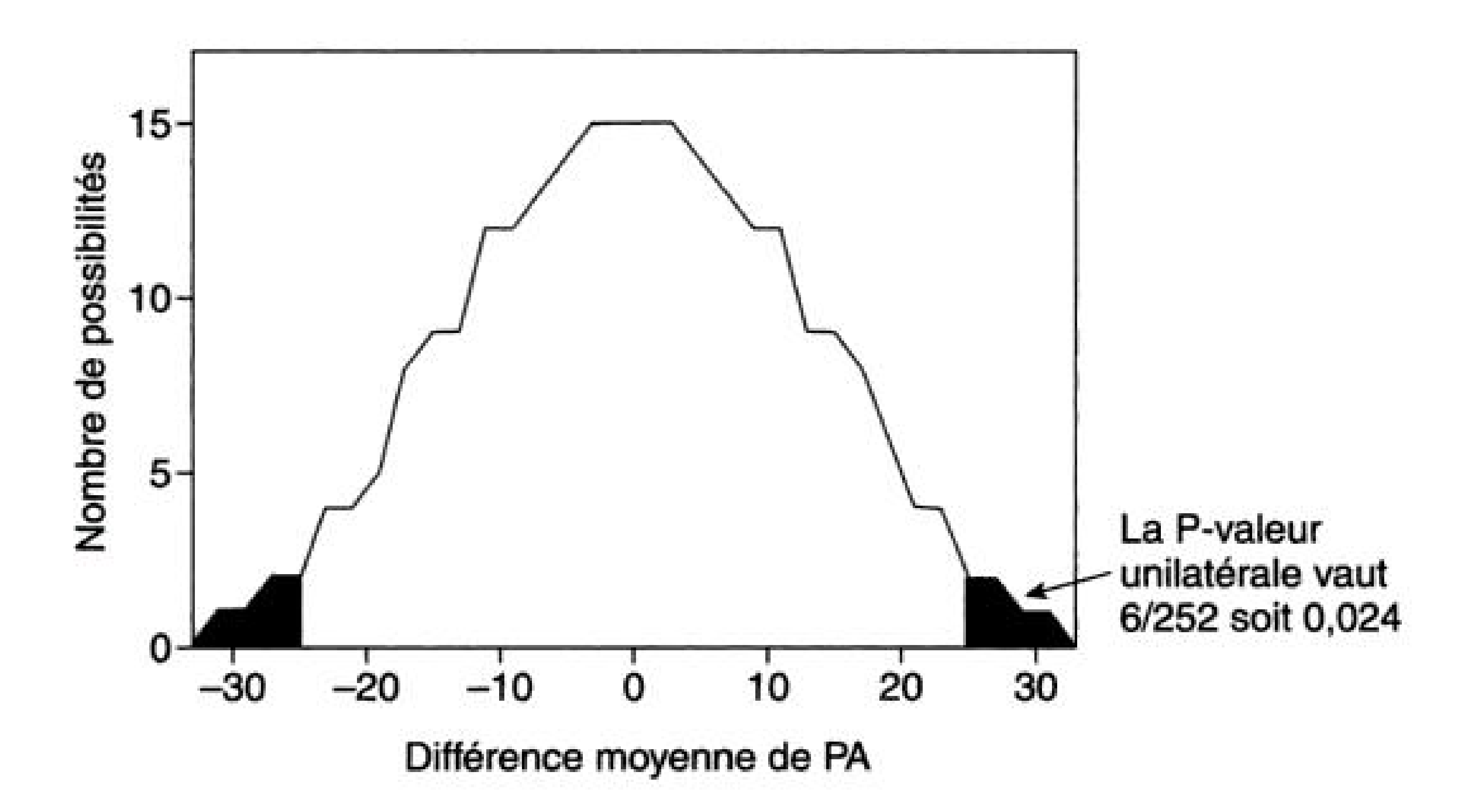


FIGURE 24.1 Test exact de randomisation. La figure représente toutes les 252 manières possibles dont les valeurs observées peuvent être réorganisées en plaçant cinq observations dans chaque groupe. En X, se trouve la différence entre les moyennes et en Y, le nombre de dispositions possibles qui conduiraient à cette différence. La surface noircie à droite représente les six dispositions qui conduisent à une différence entre les moyennes aussi grande ou plus grande que celle observée dans l'exemple. La surface noircie à gauche représente les six dispositions qui conduisent à une différence entre les moyennes aussi grande ou plus grande que celle observée dans l'exemple mais dans l'autre direction. La P-valeur bilatérale est 12/252 soit 0,048.

Il y a trois façons d'obtenir des sommes de rangs de 18 et 37 et quatre façons d'obtenir une différence plus grande entre les sommes de rangs (17 et 38 (2 fois), 16 et 39, 15 et 40). Il y a 245 autres manières d'organiser les valeurs de sorte que la différence entre les rangs soit moins extrême ou dans la direction opposée. La P-valeur unilatérale est donc $7/252 = 0{,}028$. La P-valeur bilatérale, qui intègre la possibilité que la différence entre les sommes de rangs soit aussi importante mais en faveur du groupe A, est égale à deux fois la valeur unilatérale, soit 0,056.

Parce que le test t utilise une information en plus (ou plutôt une hypothèse supplémentaire), il est plus puissant que le test de Mann-Whitney lorsque les conditions d'application du test t sont satisfaites. Avec de grands échantillons, la différence en terme de puissance est négligeable. Avec des échantillons plus petits, la différence au niveau de la puissance est plus marquée. Si l'on a 7 observations ou moins pour l'ensemble des deux groupes, le test de Mann-Whitney ne peut jamais donner une P-valeur inférieure à 0,05, aussi importante que soit la différence entre les groupes.

7 RÉALISATION DU TEST DE MANN-WHITNEY[68]

Comme pour la plupart des tests statistiques, il vaut mieux laisser les calculs aux ordinateurs. Si on souhaite calculer soi-même le test, il faut suivre les étapes décrites ci-dessous. Bien que la logique soit simple, il y a deux manières différentes de réaliser le test. La méthode décrite par Mann-Whitney implique le calcul d'une variable appelée *U*. Une autre méthode, mais équivalente, décrite par Wilcoxon implique le calcul d'une variable appelée *T*. Chaque méthode utilise un ensemble distinct de tables. Les deux méthodes sont présentées ici de façon à pouvoir être utilisées selon le type de tables disponible. Parce que Wilcoxon et Mann et Whitney ont développé des tests équivalents, on nomme parfois le test de l'un ou l'autre nom. Wilcoxon a également développé un test non paramétrique pour des données appariées que nous verrons dans la section suivante ; les tables à utiliser pour les deux tests sont très différentes mais on peut les confondre facilement.

1. Ordonner les valeurs des deux groupes. En les ordonnant, il ne faut pas tenir compte du groupe auquel appartiennent les valeurs. À la plus petite valeur pour l'ensemble des deux groupes, on attribue le rang 1 et à la plus grande valeur pour l'ensemble des deux groupes, on attribue un rang égal au nombre total d'observations des deux groupes. Si deux valeurs ou plus sont identiques, on attribue à chacune la moyenne des rangs qui leur correspondraient. Par exemple, si les valeurs du 5ᵉ et 6ᵉ rang sont identiques, on attribue à chacune un rang de 5,5. Si les trois valeurs de rang 11, 12, 13 sont identiques, on attribue à chacune le rang 12. Il faut ordonner en fonction des valeurs observées, pas des valeurs absolues. Les nombres négatifs ont toujours des rangs inférieurs à ceux des nombres positifs.
2. Sommer les rangs dans chaque groupe. On dénomme la somme des rangs T_a et T_b. Il faut regarder dans la table appropriée pour trouver la P-valeur. Bien que cette table se trouve habituellement sous l'intitulé test de « Mann-Whitney » ou test de la

[68] Cette partie contient les équations nécessaires pour calculer soi-même les statistiques. Elle peut être passée sans perdre le fil de l'exposé.

Tableau 25.1 Données de pression artérielle pour cinq étudiants

	EM1	EM2	Différence
Étudiant A	120	115	−5
Étudiant B	80	95	15
Étudiant C	90	105	15
Étudiant D	110	120	10
Étudiant E	95	100	5

Pour poursuivre l'exemple de la pression artérielle, supposons que nous suivons cinq étudiants et mesurons leur pression artérielle à la fois en première et seconde année et que nous avons recueilli les données présentées dans le tableau 25.1. La première colonne est la même que précédemment ; la deuxième colonne est différente.

Chaque ligne représente un seul étudiant dont la pression artérielle a été mesurée deux fois à un an de distance. Dans la dernière colonne se trouve l'accroissement de pression artérielle pour chaque étudiant. L'augmentation moyenne vaut 8 et l'ESM de cette augmentation est 3,74. Le test t vaut donc 2,14. Comme nous avons cinq sujets, le nombre de dl est quatre. La P-valeur bilatérale vaut 0,0993 ; l'IC à 95 % pour la variation moyenne de pression artérielle va de −2,3 à 18,4. Avec aussi peu de sujets, l'IC est très large.

Si ces données avaient été analysées avec un test t non pairé, la P-valeur bilatérale aurait été 0,370. En faisant des mesures répétées chez les mêmes sujets et en utilisant cette information dans l'analyse, l'étude est plus puissante et génère une P-valeur plus petite et un IC plus étroit.

3 CONDITIONS D'APPLICATIONS DU TEST t PAIRÉ

- Les paires doivent avoir été sélectionnées de manière aléatoire ou au minimum être représentatives d'une population plus grande.
- Les échantillons doivent être appariés. L'appariement entre les valeurs de la colonne A et celles de la colonne B doit avoir été fait sur base d'un protocole expérimental et décidé avant que les données ne soient récoltées. Il n'est pas indiqué d'apparier des données après qu'elles aient été récoltées.
- Chaque paire doit être sélectionnée de manière indépendante des autres.
- La distribution des différences dans la population doit être approximativement gaussienne.

4 TEST t « RAPPORT »

Le test t pairé est habituellement effectué de la manière décrite ci-dessus en calculant la différence entre les membres de la paires (par soustraction). L'hypothèse alternative

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

26

LA COMPARAISON DES DÉNOMBREMENTS OBSERVÉS ET ATTENDUS

1 L'ANALYSE DES DÉNOMBREMENTS

Supposons qu'en moyenne, 10 % des patients décèdent pendant ou immédiatement après une opération à risque. Cependant, au cours du mois dernier, on a observé que 16 patients sur 75 étaient décédés. On voudrait savoir si cela reflète réellement un changement ou s'il s'agit seulement d'une coïncidence. Des calculs statistiques ne peuvent donner une réponse définitive à cette question, mais ils peuvent répondre à une question liée à la précédente : si la probabilité de mourir était restée à 10 %, quelle serait la probabilité d'observer 16 décès ou plus sur 75 patients ? Si la probabilité de mourir était restée à 10 %, on s'attendrait à avoir $10\% \times 75 = 7{,}5$ décès en moyenne dans un échantillon de 75 patients. Mais dans un échantillon particulier de 75 patients, on pourrait en observer plus ou moins que le nombre attendu. Les données peuvent être résumées comme le montre le tableau 26.1.

Remarquons que ce tableau n'est pas un tableau de contingence. Dans un tableau de contingence, les colonnes doivent correspondre à des catégories différentes (par ex., homme/femme ou décédé/vivant). Ici, le tableau présente des nombres observés et attendus qui ne sont pas des issues alternatives.

L'hypothèse nulle postule que les données observées sont échantillonnées des populations sur la base des fréquences attendues. Il faut combiner les écarts entre observés et attendus et calculer ensuite une P-valeur qui répond à cette question : si l'hypothèse nulle était vraie, quelle serait la probabilité de sélectionner aléatoirement des sujets pour lesquels on observe un tel écart entre les nombres observés et attendus ?

n'a pas accès à un programme qui réalise le test de Fisher, on peut plutôt utiliser le test chi-carré. Si on utilise le chi-carré, il faut décider d'appliquer ou non la correction de Yates. Des statisticiens différents donnent des recommandations différentes et il n'y a pas de consensus ferme. Dans tous les cas (avec ou sans la correction de Yates), le test chi-carré fournit une P-valeur approximative. Plutôt que de décider quelle estimation est la meilleure en fonction de l'objectif poursuivi, il vaut mieux utiliser un programme qui calcule la P-valeur exacte avec le test de Fisher.

Grands échantillons. Si l'échantillon est de grande taille (milliers de sujets), on peut utiliser le chi-carré. Avec d'aussi grands échantillons, le calcul du test de Fisher prend du temps, peut conduire à des problèmes de calcul (dépassement de la capacité de l'ordinateur) et il n'est pas plus précis que le chi-carré. Avec de grands échantillons, cela n'a aucune importance d'utiliser ou non la correction de Yates car la valeur du chi-carré et la P-valeur seront pratiquement identiques avec ou sans correction.

Le chi-carré et le test de Fisher sont des méthodes alternatives de calcul. Les données traitées par les deux tests sont exactement les mêmes (une table 2×2) et la P-valeur résultante a la même signification. Le test de Fisher donne toujours une P-valeur théoriquement correcte mais calculer un test de Fisher sans ordinateur est plutôt fastidieux même avec des tables. Par contre, le chi-carré est facile à calculer mais la P-valeur résultante est seulement une approximation. Lorsque le nombre de sujets est suffisamment grand, l'approximation est utile. Lorsque le nombre de sujets est petit, l'approximation n'est pas valable et peut conduire à des résultats erronés.

En lisant la littérature, on peut rencontrer une troisième manière de calculer une P-valeur, le *test z*. Bien que le test z semble être tout à fait différent du chi-carré, les deux tests sont mathématiquement équivalents et conduisent à des P-valeurs identiques. Comme pour le chi-carré, il faut décider ou non d'appliquer la correction de Yates pour calculer le test z.

6 CALCUL DE LA PUISSANCE[73]

Lorsque l'on interprète les résultats d'une étude conduisant à une conclusion non significative, il est utile de calculer la puissance de cette étude. Rappelons-nous la définition de la puissance.

Si l'on suppose que les proportions vraies dans les deux populations sont π_1 et π_2, la puissance est la probabilité que l'on obtienne une différence significative entre les proportions lorsque l'on compare des échantillons aléatoires d'une certaine taille. Dans une étude prospective ou expérimentale, π_1 et π_2 sont les incidences dans les deux populations ; dans une étude cas-témoins, π_1 et π_2 sont les proportions de cas et de témoins exposés au facteur de risque ; dans une étude transversale, π_1 et π_2 sont les prévalences dans les deux populations.

La puissance d'une étude dépend des suppositions que l'on fait pour les proportions π_1 et π_2 dans les deux populations. Toutes les études ont une puissance faible si π_1 et

[73] Cette partie contient les équations pour faire soi-même le calcul des statistiques. Elle peut être sautée sans risque de perdre le fil de l'exposé.
NDT : la proportion dans la population a été notée π

Avec l'information que nous avons reçue jusqu'à présent, nous serons seulement capables de comprendre environ la moitié des articles de la littérature médicale. Les autres articles utilisent des tests statistiques plus avancés. Pour maîtriser ces tests, nous aurions besoin d'un cours de statistique approfondie ou nous devrions étudier un texte plus poussé. Les chapitres de cette partie nous donnent une vue d'ensemble sur l'utilisation de ces tests, ce qui peut parfois suffire à en comprendre les résultats. On peut lire ces chapitres dans un ordre quelconque sans perdre le fil de l'exposé.

4.1 Exemple 28.1

Nous avons placé un échantillon radioactif dans un compteur de rayons gamma et nous dénombrons les émissions radioactives pendant 10 minutes. Le compteur relève 225 coups par minute. Quel est l'intervalle de confiance à 95 % ? Puisqu'on a compté les coups pendant 10 minutes, l'instrument doit avoir détecté 2250 désintégrations radioactives. L'intervalle de confiance à 95 % de cette valeur va de 2157 à 2343 pour 10 minutes. Donc, l'intervalle de confiance à 95 % du nombre moyen de coups par minute va de 216 à 234.

4.2 Exemple 28.2

L'exposition à une toxine environnementale cause 1,6 décès pour 1000 personne-années d'exposition. Quel est l'intervalle de confiance à 95 % ? Pour le calculer, nous devons connaître le nombre exact de décès qui ont été observés dans l'étude. Dans cette étude, 16 décès ont été observés pour 10 000 personne-années (les auteurs ont peut-être étudié 10 000 personnes pendant une année ou 500 personnes pendant 20 ans). En utilisant la table A5.9 de l'annexe, on trouve un intervalle de confiance à 95 % du nombre de décès allant de 9,15 à 25,98. En divisant par 10, on trouve que l'intervalle de confiance à 95 % du taux de décès va de 0,92 à 2,6 décès pour 1000 personne-années d'exposition.

OBJECTIFS

1. Vous devez être capables d'identifier des variables qui sont distribuées selon une distribution de Poisson. En utilisant une calculatrice et un livre, vous devez être capables de calculer l'intervalle de confiance à 95 % de la moyenne d'une variable de Poisson.

EXERCICES

1. Vous utilisez un hémocytomètre pour dénombrer des globules blancs. Vous regardez une lame quadrillée au microscope et observez 25 carrés de 0,1 microlitre. Vous dénombrez les globules blancs dans 9 carrés et vous en trouvez 50 au total. Pouvez-vous calculer l'intervalle de confiance à 95 % du nombre de globules blancs par microlitre ? Quelles hypothèses faites-vous ?
2. En 1998, une publication dans *Nature* (333:816, 1988) a fait sensation dans la presse populaire et scientifique. Les auteurs prétendaient que des anticorps, même dilués à une concentration de 10^{-120} par rapport à la concentration de départ, stimulaient une dégranulation des basophiles. À une telle dilution, la probabilité qu'il y ait encore dans le tube ne fût-ce qu'une seule molécule d'anticorps est quasiment nulle. Les chercheurs ont émis l'hypothèse que l'eau « se rappelait » en quelque sorte qu'elle avait vu des anticorps. Ces résultats donnaient de la

- *Échantillonnage aléatoire simple.* Les données doivent être choisies au hasard, ou du moins être représentatives d'une population plus grande.
- *Les données doivent permettre de construire un tableau de contingence.* Les valeurs doivent correspondre au nombre de sujets. Les catégories qui définissent les lignes doivent être mutuellement exclusives ainsi que celles qui définissent les colonnes. Beaucoup de tableaux de données ne sont pas des tableaux de contingence et ne doivent pas être analysés avec le test du chi-carré.
- Le test du chi-carré est basé sur plusieurs hypothèses simplificatrices qui ne sont *valides que dans les grands échantillons.* Quel nombre peut-on considérer comme grand ? Toute règle est quelque part arbitraire mais il existe une règle communément admise pour décider s'il est correct d'utiliser le test du chi-carré. Plus de 80 % des valeurs attendues doivent être supérieures ou égales à 5 et toutes les valeurs attendues doivent être supérieures ou égales à 2 (supérieures ou égales à 1 si le tableau a plus de 30 degrés de liberté). Cette règle concerne uniquement les valeurs attendues, pas les valeurs observées. Si la règle n'est pas satisfaite, on peut combiner deux lignes ou plus, ou bien deux colonnes ou plus, tout en respectant le sens des données. Sinon, on peut utiliser un programme qui calcule un test exact (analogue au test exact de Fisher).
- *Indépendance des observations.* Chaque sujet doit avoir été sélectionné de façon indépendante dans la population.

Par exemple, le tableau 29.4 est un tableau de contingence contenant des données de pression artérielle dans 4 classes.

Les résultats de InStat sont présentés dans le tableau 29.5. La P-valeur est 0,0908. L'hypothèse nulle est que la prévalence de l'hypertension est identique dans les quatre classes. Si cette hypothèse nulle était vraie, il y aurait 9,1 % de chance de sélectionner au hasard des sujets présentant une variation aussi importante (ou plus importante) que celle observée ici.

TABLEAU 29.4 Tableau de contingence pour les données de pression artérielle

	Pression artérielle élevée	Pression artérielle non élevée
I	5	25
II	11	19
III	12	19
IV	14	16

TABLEAU 29.5 Output du logiciel Instat pour le test du chi-carré

Chi-carré = 6,472. Degrés de liberté : 3.
Taille de la table : 4 lignes, 2 colonnes.
La P-valeur est 0,0908.
Les lignes et les colonnes ne sont pas associées significativement.

2 ANOVA À UNE VOIE

L'ANOVA à une voie teste l'hypothèse nulle que toutes les populations ont une moyenne identique. Il en résulte une P-valeur qui répond à la question suivante : si l'hypothèse nulle est vraie, quelle est la probabilité que des moyennes d'échantillons sélectionnés au hasard dans une même population varient autant (ou encore plus) que ce qui a été vraiment observé ?

Comme le nom l'indique, l'ANOVA analyse la variance entre les valeurs. Rappelons-nous que la première étape du calcul d'une variance consiste à additionner les carrés des différences entre chaque valeur et leur moyenne. Cette somme s'appelle la *somme des carrés*. La variance est la moyenne des carrés et la déviation standard est la racine carrée de la variance.

Lorsqu'on combine des données relatives à plusieurs groupes, la variance a deux composantes. Si les groupes ont des moyennes différentes, une partie de la variation vient de la différence *entre* les moyennes des groupes. L'autre partie de la variation vient des différences entre les sujets *à l'intérieur* de chaque groupe. Cette autre partie peut être quantifiée en sommant les carrés des différences entre chaque valeur et la moyenne de son groupe. Cette quantité s'appelle *la somme des carrés intra-groupes*. La variation totale peut être quantifiée en sommant les carrés des différences entre chaque valeur et la moyenne *générale*. Cette quantité s'appelle *la somme des carrés totale*. Si l'hypothèse nulle est vraie, la somme des carrés intra-groupes aura une valeur proche de la somme des carrés totale parce que chaque valeur sera presque aussi proche de la moyenne générale qu'elle ne l'est de la moyenne de son propre groupe. Si l'hypothèse nulle est fausse, chaque valeur aura tendance à être plus proche de la moyenne de son propre groupe qu'elle ne l'est de la moyenne générale ; on s'attend à ce que la somme des carrés intra-groupes soit sensiblement plus petite que la somme des carrés totale. L'ANOVA à une voie calcule une P-valeur pour répondre à la question suivante : si l'hypothèse nulle était vraie, quelle serait la probabilité qu'une étude de cette taille produise une différence aussi importante (ou plus importante) entre la somme des carrés totale et la somme des carrés intra-groupes ?

Des livres entiers ont été écrits sur l'ANOVA et la description qui précède donne seulement une petite idée du fonctionnement de la méthode. La façon la plus simple de comprendre les résultats d'une ANOVA est de suivre un exemple.

3 UN EXEMPLE

Continuons l'exemple 13.5 de la page 150, dont les résultats sont repris dans le tableau 30.1.

Il vaut mieux laisser les calculs aux ordinateurs. Les équations ne seront pas présentées ici. Les résultats obtenus avec le logiciel InStat sont présentés dans le tableau 30.2. D'autres logiciels fourniraient des résultats analogues, à l'exception de quelques variantes de vocabulaire. En français par exemple, certains programmes pourraient utiliser le mot *modèle* à la place de l'expression *entre les groupes (traitements)*, le mot *erreur* à la place de *résidus* et l'expression ANOVA à *un facteur* ou ANOVA *à un critère* à la place d'ANOVA

multiples. La plupart des programmes informatiques offrent une panoplie de tests post-hoc et en choisir un est quelque peu déconcertant. Il peut être utile de considérer la réponse aux questions suivantes.

- Les colonnes suivent-elles un ordre naturel, comme un ordre chronologique ou des doses croissantes ? Si oui, il faut considérer les tests post-hoc de tendance. Au lieu de comparer des moyennes deux à deux, le test post-hoc de tendance regarde s'il existe une tendance linéaire significative dans les moyennes selon l'ordre des groupes. C'est en fait une forme de régression linéaire.
- Souhaitons-nous faire des comparaisons élaborées ? Par exemple, on pourrait vouloir comparer la moyenne de tous les groupes traités avec la moyenne d'un groupe témoin. Ou, on pourrait vouloir comparer la moyenne des groupes A et B avec celle des groupes C, D et E. Les statisticiens appellent de telles comparaisons des *contrastes*. Pour calculer des contrastes, il faut utiliser la méthode de Scheffé. Pour prendre en compte la grande variété de contrastes possibles, la méthode de Scheffé génère des IC qui sont plus larges. Elle a aussi moins de puissance statistique pour détecter des différences que d'autres tests post-hoc.
- Souhaitons-nous comparer un groupe témoin à tous les autres groupes (sans comparer les autres groupes entre eux) ? Si oui, il faut choisir le test de Dunnett.
- En se basant sur le design expérimental, est-il sensé de ne comparer que certaines paires précises de moyennes ? Si oui, il faut choisir le test de Bonferroni pour les paires de moyennes à comparer mais le choix de ces paires doit se faire sur base du protocole expérimental. Il n'est pas correct de regarder d'abord les données et de décider ensuite des paires à comparer. En regardant d'abord les données, on compare implicitement toutes les colonnes et cela doit se faire explicitement.
- Souhaitons-nous comparer toutes les moyennes deux à deux ? Trois méthodes sont couramment utilisées : celle de Bonferroni, celle de Tukey et celle de Student-Newman-Keuls. Le test de Bonferroni est bien connu et relativement facile à comprendre, mais il génère des IC qui sont trop larges et sa puissance est trop faible. Avec trois ou quatre groupes, on ne remarquera pas la différence mais il ne

TABLEAU 30.3 Résultats de InStat pour le test post-hoc de Tukey

Comparaison	Différence moyenne	q	P-valeur
Non coureuse vs amateur	0.1400	2.741	** $P < 0.01$
Non coureuse vs élite	0.1200	2.741	ns $P > 0.05$
Amateur vs élite	−0.02000	0.4574	ns $P > 0.05$

Différence	Différence moyenne	Limite inférieure de l'IC à 95 %	Limite supérieure de l'IC à 95 %
Non coureuse - amateur	0.1400	0.03823	0.2418
Non coureuse - élite	0.1200	−0.02688	0.2669
Amateur - élite	−0.02000	−0.1667	0.1267

valeur de ces coefficients : si X_j change d'une unité alors que toutes les valeurs des autres X sont constantes, alors Y augmente de β_j (Y décroît si β_j est négatif). Dans l'exemple, le coefficient β de la variable X « traitement diurétique antérieur » valait −8,8 ml/min. Ceci signifie qu'après ajustement pour toutes les autres variables, les sujets qui avaient pris des diurétiques dans le passé, avaient une clearance de la créatinine moyenne, 8,8 ml/min plus basse que celle des sujets qui n'avaient pas pris de diurétiques. Le coefficient β du log (plomb) valait −9,5 ml/min. Ceci signifie qu'après ajustement pour toutes les autres variables, une augmentation de 1 unité en log(plomb) (ce qui signifie que la concentration en plomb à été multipliée par 10) est associée à une diminution de la clearance de la créatinine de 9,5 ml/min. L'IC à 95 % s'étendait de −18,1 à −0,9 ml/min.

Les programmes de régression multiple calculent également R^2 qui est la fraction de la variabilité[86] totale en Y « expliquée » par la variation en toutes les variables X[87]. Le programme génère une valeur de R^2 pour la régression dans son ensemble, pas une pour chaque X. Dans l'exemple, $R^2 = 0,27$. Ceci signifie que 27 % de la variabilité de la clearance de la créatinine était expliquée par la variabilité en plomb, âge et toutes les autres variables. Il reste donc 73 % de variabilité qui n'est pas expliquée par la variabilité en l'une quelconque des variables mesurées. De nombreux programmes fournissent aussi un R^2 ajusté. L'ajustement diminue la valeur en fonction du nombre de variables X. Le R^2 ajusté est une meilleure mesure d'ajustement du modèle que la valeur non ajustée.

Les programmes de régression multiple calculent plusieurs P-valeurs. Une P-valeur teste l'hypothèse globale que tous les coefficients β valent 0 dans la population. En d'autres mots, c'est le test de l'hypothèse nulle qu'aucune des variables X n'influence Y. Si cette P-valeur est basse, on peut rejeter l'hypothèse nulle globale qu'aucune des variables X n'influence la variable étudiée. Dans l'exemple, les chercheurs n'ont pas fourni la P-valeur globale. Le programme fournit des P-valeurs individuelles pour chaque coefficient β_j, testant l'hypothèse nulle que ce coefficient spécifique β_j vaut 0, ce qui est nettement plus intéressant. Exprimé autrement, chacune de ces hypothèses nulles postule que la variable X concernée n'est pas associée linéairement à Y, après ajustement pour toutes les autres variables X. Dans l'exemple, les auteurs ont spécifié que la P-valeur pour le plomb était $< 0,05$. Si réellement, il n'y avait pas d'association entre la concentration en plomb et la clearance de la créatinine, on trouverait une corrélation aussi importante dans une étude de cette taille (après ajustement pour l'âge et les autres variables) moins de 5 % du temps.

La régression multiple est une technique puissante mais il peut être difficile d'interpréter les résultats. Un des problèmes est la difficulté de visualiser les résultats graphiquement. Ci-dessous sont énumérés quelques autres problèmes qui peuvent compliquer l'interprétation des résultats de la régression multiple :

- *Valeurs extrêmes*. Comme dans la régression simple, un seul point extrême peut distordre considérablement les résultats. Dans le cas de la régression simple, ceci apparaît généralement de façon évidente lorsqu'on regarde le graphique des données. Il est plus difficile de visualiser graphiquement les données de régression

[86] NDT : dans l'interprétation du R^2, « variance » a été traduit par « variabilité ».

[87] Dans le cadre de la régression multiple, la tradition est d'utiliser un R majuscule. Dans le cadre de la régression simple, la tradition est d'utiliser un r minuscule.

- Pas d'interaction. L'influence d'une variable X particulière quelconque est la même pour toutes les valeurs des autres variables X. Dans l'exemple, le modèle postule qu'une fraction d'éjection basse augmenterait le risque de sujets jeunes ou âgés de façon équivalente.

4 L'INTERPRÉTATION DES RÉSULTATS DE LA RÉGRESSION LOGISTIQUE

Le programme de régression logistique fournit les estimations de meilleur ajustement soit pour chaque rapport des cotes (équation 32.2), soit pour chaque coefficient β (équation 32.1). Si on lit un article (ou un output d'ordinateur) qui utilise l'équation 32.1, il faut convertir les coefficients β en rapports des cotes de la manière suivante : le rapport des cotes pour la j^e^ variable est égal à e^{β_j}.

Lorsque X_j est une variable binaire (par ex. X = 0 homme, X = 1 femme), l'interprétation du rapport des cotes est celle qui nous est familière. Toutes les autres variables étant constantes, le rapport des cotes est la cote de l'événement pour quelqu'un qui a X = 1 divisée par la cote de l'événement pour quelqu'un qui a X = 0. Dans l'exemple, les auteurs ont codé une des variables X en X = 0 si le sujet n'avait pas une fraction d'éjection basse et X = 1 s'il avait une fraction d'éjection basse. Le rapport des cotes est la cote de complication cardiaque parmi les patients ayant une fraction d'éjection basse divisée par la cote de complication cardiaque parmi les patients avec une fraction d'éjection normale. C'est un rapport des cotes ajusté dans la mesure où il est corrigé pour l'influence de toutes les autres variables X telles que l'âge.

Lorsque X_j est une mesure, l'interprétation du rapport des cotes nous est moins familière. Le rapport des cotes est l'accroissement relatif de la cote lorsque X_j augmente de 1,0. Toutes les autres variables X étant constantes, le rapport des cotes est la cote de l'événement pour quelqu'un qui a $X_j = Z$ (où Z représente n'importe quelle valeur) divisée par la cote de l'événement pour quelqu'un qui a $X_j = Z - 1$. Avec cette méthode, le rapport des cotes est supposé être le même pour toutes les valeurs de Z.

Remarquons que les programmes de régression logistique ne font pas la distinction entre les variables auxquelles on s'intéresse particulièrement (dans l'exemple, la diminution de la fraction d'éjection) et les variables pour lesquelles on ajuste (dans l'exemple, l'âge et les antécédents de maladie coronarienne). La régression logistique conduit à des rapports des cotes pour chaque variable X, ajustés pour toutes les autres. Dans l'exemple, le programme fournirait également un rapport des cotes ajusté pour l'augmentation du risque de complication cardiaque avec l'augmentation de l'âge et les antécédents de maladie coronarienne. C'est seulement en interprétant les résultats que l'on peut distinguer les variables auxquelles on s'intéresse des variables pour lesquelles on ajuste.

Les programmes de régression logistique fournissent plusieurs P-valeurs. Une P-valeur teste l'hypothèse nulle que le rapport des cotes global dans la population vaut 1,0. En d'autres mots, cette hypothèse nulle postule qu'aucune des variables X n'est associée à l'événement. Si cette P-valeur est basse, on peut considérer les P-valeurs individuelles de chaque rapport des cotes. Chacune de ces P-valeurs teste l'hypothèse nulle que le

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

et sont allés en Oklahoma, ils ont relevé le niveau d'intelligence moyen dans les deux états. »)[94].

5 LA RÉGRESSION MULTIPLE AVEC DES DONNÉES DE SURVIE : RÉGRESSION DES RISQUES INSTANTANÉS PROPORTIONNELS

La régression des risques instantanés proportionnels applique la méthodologie de la régression aux données de survie. Cette méthode permet de comparer la survie de deux groupes ou plus après ajustement pour les autres variables.

5.1 Exemple 33.1 Suite (Diazépam et accès fébriles)

Les chercheurs ont appliqué le modèle de régression des risques instantanés proportionnels pour ajuster l'effet du traitement en fonction de l'âge, du nombre d'accès fébriles antérieurs et de plusieurs autres variables. Après avoir effectué ces ajustements, ils ont obtenu un risque relatif de 0,61 avec un IC à 95 % s'étendant de 0,39 à 0,94. Les sujets traités au diazépam avaient un risque d'accès fébrile qui valait seulement 61 % du risque des sujets traités avec le placebo. Cette réduction était statistiquement significative avec une P-valeur de 0,027. Si le diazépam n'était pas efficace, il y aurait seulement 2,7 % de chance d'observer un risque relatif aussi bas dans une étude de cette taille. Cet exemple montre que les résultats du modèle de régression des risques instantanés proportionnels sont faciles à interpréter même si les détails de l'analyse sont complexes.

6 COMMENT LE MODÈLE DE RÉGRESSION DES RISQUES INSTANTANÉS PROPORTIONNELS FONCTIONNE

Une courbe de survie représente graphiquement la proportion de survivants en fonction du temps. La pente ou la dérivée de la courbe de survie est le taux de mortalité pendant un court intervalle de temps. Ce taux est appelé *risque instantané* ou *hazard* en anglais. Par exemple, si l'on s'attend à ce que 20 % des patients avec un certain type de cancer meurent cette année, alors le risque instantané est de 20 % par an. En comparant deux groupes, les chercheurs supposent souvent que le rapport des risques instantanés de décès est constant au cours du temps. Par exemple, le risque instantané parmi les patients traités pourrait être la moitié de celui des patients témoins. Le taux de mortalité change au cours de l'étude, mais à n'importe quel moment, le risque de décès des patients traités vaut la moitié de celui des patients témoins. Une autre manière d'exprimer cela est de dire que les deux fonctions de risque instantané sont proportionnelles l'une par rapport à l'autre. C'est une hypothèse raisonnable pour de nombreuses situations cliniques.

---

[94] AR Feinstein, DA Sosin, CK Wells. Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *New Engl J Med* 312 : 1604-1608.

tration augmente encore, la liaison commence à atteindre un plateau lorsqu'on approche 100 % de saturation de récepteurs par le ligand.

Le radioligand se fixe façon réversible à de nombreux types de récepteurs :

$$\text{Récepteur} + \text{Ligand} \overset{K_d}{\rightleftharpoons} (\text{Récepteur} \cdot \text{Ligand}) \tag{34.1}$$

À partir de la définition chimique de l'équilibre et de la loi d'action de masse, la relation entre la concentration de radioligand et la quantité de liaison à l'équilibre suit l'équation 34.2 :

$$\text{Liaison spécifique} = \frac{B_{max} \cdot [\text{Ligand}]}{K_d + [\text{Ligand}]}$$
$$Y = \frac{B_{max} \cdot X}{K_d + X} \tag{34.2}$$

La relation entre X (concentration) et Y (liaison) n'est pas linéaire, aussi les méthodes de régression linéaire ne peuvent être utilisées. Nous devons faire passer une courbe à travers les observations pour trouver les valeurs de meilleur ajustement de B_{max} et K_d.

3 QUEL EST LE PROBLÈME AVEC LA TRANSFORMATION DE DONNÉES CURVILIGNES EN DROITE ?

Avant la période des ordinateurs bon marché, la régression non linéaire n'était pas facilement accessible à la plupart des chercheurs. Plutôt que de l'utiliser, les chercheurs transformaient leurs données de manière à rendre le graphique linéaire. Ils utilisaient ensuite la régression linéaire pour analyser les données transformées. Pour les données de type particulier envisagées dans l'exemple, la transformation de Scatchard linéarise les données, ainsi qu'on peut le voir dans la figure 34.2.

D'autres transformations linéaires couramment utilisées, sont le double réciproque ou le diagramme de Lineweaver-Burke utilisé en enzymologie et la transformation logarithmique utilisée dans les études cinétiques. Toutes ces méthodes sont dépassées et devraient être très peu utilisées pour analyser des données. En effet, la régression linéaire suppose que la dispersion des données autour de la droite suit une distribution gaussienne avec une déviation standard (DS) qui ne dépend pas de la valeur de X. Cette hypothèse est rarement vérifiée avec des données transformées. De plus la régression linéaire suppose que X et Y sont mesurés de façon indépendante. Avec certaines transformations (par exemple, la transformation de Scatchard), les valeurs de X et Y sont permutées par la transformation, ce qui rend la régression linéaire non valide.

La figure 34.3 illustre le problème lié à la transformation des données. La partie gauche présente des données qui proviennent d'une expérience similaire à celle de l'exemple précédent mais la dispersion est plus grande. La courbe en trait plein a été ajustée par régression non linéaire. La partie de droite présente les mêmes données après la transformation de Scatchard - c'est le diagramme de Scatchard. La transformation de Scatchard

35

LA COMBINAISON DE PROBABILITÉS

1 PROPAGATION DES ERREURS

Les variables intéressantes biologiquement ou cliniquement ne sont pas souvent mesurées directement mais sont plutôt calculées à partir de 2 ou plusieurs variables mesurées. Ce chapitre explique comment déterminer l'intervalle de confiance (IC) de la variable calculée.

1.1 L'intervalle de confiance d'une différence

Il est fréquent que l'on soit amené à analyser une différence (D) définie en soustrayant une valeur de base ou une mesure non spécifique (B) d'une mesure totale (A). En d'autres mots, D = A – B. À la fois A et B sont des variables gaussiennes et l'on veut connaître l'intervalle de confiance à 95 % de la différence entre les moyennes (D). Si on suppose que les observations ne sont pas appariées et que les tailles des deux échantillons sont égales, on peut calculer l'erreur standard de la différence entre les moyennes, à partir de l'équation suivante :

$$ES_D^2 = \sqrt{ESM_A^2 + ESM_B^2} \tag{35.1}$$

Trois remarques :

- L'ES de la différence est plus grande à la fois que l'ES de A et l'ES de B.
- L'équation 35.1 (identique à l'équation 7.1) suppose que les deux groupes ont le même nombre d'observations. Si les tailles d'échantillons sont différentes il faut plutôt utiliser l'équation 7.2.
- A et B doivent être tous deux mesurés dans les mêmes unités, et D est mesuré dans les mêmes unités.

de chaque étude. Même si chaque étude individuellement n'arrive pas à une conclusion claire, une meta-analyse peut elle, conduire à des conclusions bien nettes.

Lorsqu'on lit le compte-rendu d'une meta-analyse, il faut se poser certaines questions : comment les auteurs ont-ils sélectionné les articles à prendre en considération ? Quelle proportion des articles ont-ils inclus dans l'analyse formelle ? Pourquoi ont-ils exclu les autres ? Qui étaient les patients dans chaque étude ? Les diagnostics et les traitements étaient-ils similaires ? Comment les résultats ont-ils été combinés ? Les auteurs ont-ils pris en considération différents sous-groupes de patients séparément ? Qui a payé l'étude ? Dans quelle mesure, les résultats des études étaient-ils similaires ?

TABLEAU 36.1 Résultats de la première étude

	Résultat			Analyse
	Échec	Réussite	Total	P = 1,00 RR = 1,00
Opération I	120	120	240	
Opération II	40	40	80	
Total	160	160	320	

TABLEAU 36.2 Résultats de la deuxième étude

	Résultat			Analyse
	Échec	Réussite	Total	P = 1,00 RR = 1,00
Opération I	20	60	80	
Opération II	40	120	160	
Total	60	180	240	

TABLEAU 36.3 Résultats combinés de la première et de la deuxième étude

	Résultat			Analyse
	Échec	Réussite	Total	P = 0,014 RR = 1,31
Opération I	140	180	320	
Opération II	80	160	240	
Total	220	340	560	

Que se passerait-il si on combinait les deux études et qu'on analysait les données groupées ? Les résultats, présentés dans le tableau 36.3, sont surprenants.

Globalement, l'opération I a échoué chez 140/320 = 43,75 % des sujets et l'opération II a échoué chez seulement 80/240 = 33,3 % des sujets. Le risque relatif vaut 1,31. Si on teste l'hypothèse nulle selon laquelle il n'existe pas de différence avec le test de Fisher, la P-valeur vaut 0,014. L'opération II est significativement meilleure. Si on interprétait ces résultats tels quels (ce n'est pas une chose à faire !), on conclurait que la différence entre les deux opérations est statistiquement significative et que l'opération II est meilleure de 31 %.

Comment est-ce possible ? Dans chacune des études, les deux opérations se valent. Pourquoi n'est ce pas le cas avec les données groupées ? Les analyses groupées ne respec-

Tableau 37.1 Le choix d'un test statistique

Objectif	Type de données			
	Mesure (provenant d'une population gaussienne)	Rang, score ou mesure (provenant d'un population non gaussienne)	Binomiale (deux résultats possibles)	Temps de survie
Décrire un groupe	Moyenne, DS	Médiane, écart interquartile	Proportion	Courbe de survie de Kaplan-Meier
Comparer un groupe à une valeur hypothétique	Test t pour un échantillon	Test de Wilcoxon	Test du chi-carré ou test binomial**	/
Comparer deux groupes non appariés	Test t non pairé	Test de Mann-Withney	Test exact de Fisher (test du chi-carré pour les grands échantillons)	Test du log-rank ou de Mantel-Haenszel*
Comparer deux groupes appariés	Test t pairé	Test de Wilcoxon	Test de Mc Nemar	Régression conditionnelle des risques instantanés proportionnels**
Comparer trois groupes indépendants ou davantage	ANOVA à une voie	Test de Kruskal-Wallis	Test du chi-carré	Régression des risques instantanés proportionnels de Cox*
Comparer trois groupes appariés ou davantage	ANOVA pour mesures répétées	Test de Friedman	Q de Cochran**	Régression conditionnelle des risques instantanés proportionnels**
Quantifier une association entre deux variables	Corrélation de Pearson	Corrélation de Spearman	Coefficients de contingence**	
Prédire une valeur à partir d'une autre variable mesurée	Régression linéaire simple ou régression non linéaire	Régression non paramétrique**	Régression logistique simple*	Régression des risques instantanés proportionnels de Cox*
Prédire une valeur à partir de plusieurs variables mesurées ou binomiales	Régression linéaire multiple* ou régression non linéaire multiple**		Régression logistique multiple*	Régression des risques instantanés proportionnels de Cox*

* Seulement brièvement mentionné dans ce livre.
** Non discuté dans ce livre.

3 IL FAUT SE MÉFIER DES COMPARAISONS MULTIPLES

Quand on analyse des données aléatoires, une comparaison sur vingt en moyenne sera statistiquement significative par hasard. Il faut se méfier des grandes études qui réalisent des douzaines ou des centaines de comparaisons puisqu'on est susceptible d'y rencontrer des résultats significatifs à tort. En lisant des articles, il faut se demander combien d'hypothèses les chercheurs ont testées.

4 IL NE FAUT PAS SE FOCALISER SUR LES MOYENNES : LES VALEURS EXTRÊMES PEUVENT ÊTRE IMPORTANTES

Les tests statistiques (test t, ANOVA) comparent des moyennes. La variabilité dans les études biologiques ou cliniques n'est pas toujours essentiellement due à des incertitudes dans la mesure. Au contraire, la variabilité des données reflète une réelle diversité biologique. Il faut apprécier cette diversité ! Ne soyons pas hypnotisés par les moyennes ; les valeurs extrêmes sont parfois plus intéressantes. Certains chercheurs ont gagné le prix Nobel grâce à des études portant sur des individus dont les valeurs étaient éloignées de la moyenne.

5 IL EST NORMAL DE RENCONTRER DES DISTRIBUTIONS NON GAUSSIENNES

De nombreux tests statistiques reposent sur l'hypothèse que les données proviennent d'une distribution gaussienne et les scientifiques semblent souvent penser que la nature est assez bonne pour leur montrer que toutes les variables intéressantes se distribuent de la sorte. Ce n'est pas vrai ! De nombreuses variables intéressantes ne suivent pas une distribution gaussienne.

6 QUALITÉ À L'ENTRÉE ÉGALE QUALITÉ À LA SORTIE

Les calculs statistiques ingénieux ne sont d'aucune utilité si les données n'ont pas été collectées de façon adéquate. Les statistiques constituent la partie facile de l'interprétation des données. La partie difficile consiste à évaluer la méthodologie de l'étude. Les tests statistiques ne permettent pas de savoir si l'étude a été conduite de façon appropriée. De nombreuses études sont mal conçues et les tests statistiques de telles études peuvent donner des résultats qui induisent en erreur. Si les données ne sont pas collectées de façon appropriée (ou comme dirait un biostatisticien, si l'étude est biaisée), les tests statistiques sont difficiles à interpréter. Réfléchir au design de l'étude est le défi de la science. Les études

1
RÉFÉRENCES

TABLEAU A1.1 Livres de référence en statistique

Sujet	Référence (voir liste)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Introduction intuitive aux principes de statistique	*						*		*				
Explication mathématique des principes de statistique					*		*				*		
Critique de l'utilisation de la statistique		*		*									
ANOVA à une voie							*	*				*	
ANOVA à deux voies								*			*	*	
Tests non paramétriques			*									*	
Régression multiple et logistique					*			*					*
Régression non linéaire								*		*			
Courbes de survie	*	*			*								
Taux et proportions					*	*							
Taille d'échantillon et puissance	*					*						*	
Algorithmes informatiques										*			
Design des études cliniques	*	*		*									
Référence détaillée	*				*							*	
Interprétation des résultats	*			*	*								
Épidémiologie					*	*							*

1.4 Système unique de vérification de l'analyse

Avec les autres logiciels statistiques, on peut facilement recevoir une réponse correcte à la mauvaise question. C'est pourquoi InStat fournit un système unique de vérification de l'analyse. Il effectue une double vérification pour s'assurer qu'avec les données, aucune condition d'application du test n'est violée et que le test choisi est en accord avec le protocole expérimental et répond réellement à la question posée.

2 GRAPHPAD PRISM®

GraphPad Prism, disponible en version Windows et Macintosh, combine des graphiques scientifiques, l'ajustement de courbe et des statistiques. En fait, tous les graphiques de ce livre ont été faits avec Prism.

2.1 Approche originale

L'entièreté du travail (toutes les données associées à l'étude, résultats, graphiques, mise en page et notes) peut être sauvegardée dans un seul fichier organisé. Tous les éléments sont liés — si l'on remplace ou édite des données, Prism remet à jour automatiquement le reste du travail.

2.2 Graphiques instantanés

Il suffit d'introduire les données et de cliquer une fois. Prism crée instantanément un graphique incluant des barres d'erreur automatiques et une légende. L'étude peut ainsi être résumée en combinant graphiques, données et résultats en une seule page.

2.3 Ajustement facile de courbes

Des courbes correspondant à différentes conditions expérimentales peuvent être ajustées en une seule étape. Il suffit de choisir une équation à partir de la liste extensive incluse dans Prism et celui-ci fait le reste automatiquement.

2.4 Aide statistique

Les écrans d'aide de Prism et le manuel aident à choisir les analyses et interpréter les résultats. Il ne faut pas être un expert en statistique pour utiliser Prism.

3 PLUS D'INFORMATION ET DÉMONSTRATION

Pour plus d'informations sur InStat et Prism et pour décharger des versions de démonstration gratuites, on peut consulter le site web de GraphPad à l'adresse : www.graphpad.com ou encore contacter GraphPad Software, Inc. :

Pour interpréter ces résultats, vous devez définir la population. S'il s'agissait de données réelles, vous souhaiteriez savoir comment les investigateurs ont sélectionné les patients. Sont-ils particulièrement malades ? Représentent-ils l'ensemble des patients atteints d'une certaine pathologie ou seulement une petite partie ? Sans information sur la méthode de sélection des patients, les données sont difficiles à interpréter.

Si vous pouvez définir la population d'où proviennent ces patients, vous devez supposer que ces 100 patients sont représentatifs de l'ensemble des patients et qu'ils ont été sélectionnés chacun de façon indépendante. Si ces hypothèses sont respectées, alors, vous pouvez être surs à 95 % que la probabilité globale de décès dans cette population se trouve quelque part entre les limites de l'IC à 95 %.

2. Un nouveau médicament est testé chez 100 patients et diminue la pression artérielle en moyenne de 6 %. Pouvez-vous calculer l'IC à 95 % de la réduction relative de la pression artérielle par ce médicament? Si oui, calculez cet intervalle. Si non, de quelles informations devez-vous disposer ? En quoi consiste l'IC de la réduction relative de la pression artérielle ? Quelles hypothèses devez-vous faire ?

   Cela ressemble à première vue à l'exercice précédent mais c'est pourtant assez différent. Dans l'exercice 1, le pourcentage de décès était vraiment une proportion. Dans l'exercice 2, le pourcentage représente en fait un changement dans une mesure. Vous ne pouvez pas utiliser l'équation 2.1. Vous apprendrez à calculer des IC pour ce genre de données dans les prochains chapitres et pour le faire vous devrez avoir une idée de la dispersion des valeurs (la déviation standard ou l'erreur standard).

   Faites attention aux pourcentages ! Les pourcentages peuvent exprimer une proportion, une différence relative de proportions, une différence relative entre deux mesures ou encore d'autres choses. N'utilisez la méthode du chapitre 2 qu'avec des proportions.

3. On étudie la viabilité de cellules par une coloration au bleu de trypan. Avec un hémocytomètre, on dénombre 94 cellules non teintées (viables) et 6 cellules teintées (indiquant qu'elles ne sont pas viables). Pouvez-vous calculer l'IC à 95 % de la proportion de cellules teintées (mortes) ? Si oui, calculez cet intervalle. Si non, de quelles informations devez-vous disposer ? Quelles hypothèses devez-vous faire ?

   Il s'agit essentiellement du même problème que dans l'exercice 1. Le tube entier de cellules constitue la population et on a évalué la viabilité d'un échantillon de ces cellules. Si le tube a été bien mélangé, il est raisonnable de penser que l'échantillon a été sélectionné de façon aléatoire dans la population. Vous pouvez calculer un IC à 95 % de la proportion de cellules viables dans l'entièreté de la population. En utilisant l'équation 2.1, vous pouvez dire que vous êtes surs à 95 % que la proportion de cellules teintées (mortes) se trouve entre 0,013 et 0,107.

4. En 1989, 20 étudiants sur 125 inscrits en seconde année de médecine à San Diego ont raté l'examen écrit de biostatistique. Pouvez-vous calculer l'IC à 95 % de la probabilité de réussir cet examen ? Si oui, calculez cet intervalle. Si non, de quelles informations devez-vous disposer ? Quelles hypothèses devez-vous faire ?

sentent l'activité enzymatique en unités par minute par milligramme de protéine membranaire.

Résumez ces données comme vous le feriez pour une publication. Les lecteurs ne s'intéressent pas aux résultats individuels de chaque jour ; une moyenne générale avec une mesure de la dispersion peut suffire. Présentez les résultats sous forme d'une moyenne, d'une valeur d'erreur et sans oublier n. Justifiez vos décisions.

Cet exercice contient deux difficultés. Premièrement, vous devez décider de ce que vous allez faire de la deuxième expérience du mardi. Le résultat est très différent des autres. Décider ce qu'il faut faire de valeurs extrêmes est un problème difficile et les scientifiques adoptent des attitudes différentes. Dans ce cas-ci, il est assez difficile de croire que cette valeur est juste. Non seulement elle est très éloignée des autres résultats du même jour mais en plus elle est très éloignée des résultats des six autres expériences réalisées les autres jours. Comme l'expérience a été menée sur des cellules en culture, il n'existe pas de variabilité biologique d'un jour à l'autre. Les seules sources de variation sont les erreurs expérimentales. Il semble donc évident que la valeur 967 n'est pas correcte. Dans un premier temps, il serait utile de consulter le cahier de notes du laboratoire pour voir si l'expérimentateur n'a rien noté de particulier à propos de ce tube en cours d'expérience. Si le chercheur a signalé un problème avec cette expérience, on pourrait l'écarter en toute bonne conscience. Mais même sans cela, on pourrait la supprimer des analyses en veillant à noter avec précision dans un livre de notes la démarche qui a été suivie. Inclure cette valeur modifierait tous les calculs et les résultats seraient dépourvus d'intérêt.

Si les lignes représentaient différents patients (plutôt que différentes expériences réalisées sur des cellules issues de la même culture), il faudrait adopter une autre attitude. Dans le cas de patients, la variabilité biologique joue un rôle. Peut-être que la valeur la plus grande est juste et que les deux petites sont fausses. Peut-être que le deuxième patient est réellement différent du premier et du troisième. Si c'est possible, il faudrait faire une seconde fois l'expérience avec le deuxième patient pour pouvoir trancher.

Le second problème consiste à regrouper les huit valeurs restantes. Comment faire ? Il n'est pas tout à fait juste de les considérer comme des mesures indépendantes. Les expériences réalisées le même jour se ressemblent plus qu'elles ne ressemblent aux expériences réalisées les autres jours (la variabilité est d'ailleurs plus importante d'un jour à l'autre que d'une expérience à l'autre). Par conséquent, il n'est sans doute pas valide de calculer la moyenne et la DS ou l'ESM des huit valeurs.

Par contre, on peut calculer la valeur moyenne de chaque jour, calculer ensuite la moyenne et l'ESM des trois moyennes et enfin l'IC à 95 %. Cela a du sens si on considère que la population est constituée des moyennes de toutes les expériences avec n = 3. Nous pouvons être sûrs à 95 % que la vraie moyenne se trouve dans l'intervalle de confiance[103].

---

[103] Il serait mieux d'utiliser une technique qui tiendrait également compte de la variabilité entre les expériences. Mais c'est difficile à calculer, confus à expliquer et cela ne changerait pas beaucoup les résultats.

6 CHAPITRE 7

1. (Même données que celles de l'exercice 4, chapitre 5). Le niveau sérique d'une hormone (le facteur Y) vaut 93 ± 1,5 (moyenne ± ESM) chez 100 femmes non enceintes et 110 ± 2,3 (moyenne ± ESM) chez 100 femmes au premier trimestre de leur grossesse.

   a) Que vaut l'IC à 95 % de la différence entre les niveaux moyens du facteur Y ?

   L'ES d'une différence est la racine carrée de la somme des carrés des valeurs des deux ESM (équation 7.1) soit 2,75 unités/ml dans cet exercice. Avec 100 femmes dans chaque groupe, il y a 100 + 100 − 2 = 198 degrés de liberté. Pour autant de degrés de liberté, $t^* = 1{,}97$. La différence entre les moyennes vaut 110 − 93 = 17. L'IC à 95 % s'étend à 1,97 ES de part et d'autre soit de 11,6 à 22,4 unités/ml. Si nous supposons que les deux échantillons sont représentatifs des populations, nous pouvons être sûrs à 95 % que le niveau moyen du facteur Y chez les femmes enceintes est 11,6 à 22,4 unités/ml plus élevé que chez les femmes non enceintes.

   b) Quelles hypothèses devez-vous faire pour répondre à cette question ?

   Nous faisons l'hypothèse que les deux échantillons sont représentatifs des populations de femmes enceintes et non enceintes des mêmes tranches d'âge. Puisque les échantillons sont d'assez grande taille, nous n'avons pas besoin de faire l'hypothèse que les distributions sont gaussiennes. Nous devons supposer que les deux populations ont les mêmes déviations standards et que les femmes ont chacune été sélectionnées de façon indépendante.

2. Pullan et coll. ont étudié l'utilisation de nicotine transdermique pour traiter la colite ulcéreuse. Le niveau de nicotine plasmatique de départ valait 0,5 ± 1,1 ng/ml (moyenne ± DS ; n = 35). Après 6 semaines de traitement, le niveau plasmatique valait 8,2 ± 7,1 ng/ml (n = 30).

   Calculez l'IC à 95 % de l'augmentation du niveau de nicotine plasmatique.

   L'IC à 95 % de la différence va de 5,3 à 10,1 ng/ml (sur base des équations 7.2 et 7.3).

   Quelles hypothèses faites-vous ? Sont-elles respectées ?

- Les sujets sont représentatifs de tous les patients avec une colite ulcéreuse. Cette hypothèse paraît raisonnable mais si voulez en être sûrs, vous devez vous renseigner sur la façon dont les investigateurs ont recruté les sujets.
- Les populations suivent une distribution gaussienne. Comme les échantillons sont assez grands, ce n'est pas très important. Sur base des données, il est évident que les valeurs ne suivent pas une distribution gaussienne. Si les populations étaient gaussiennes, 95 % des données seraient comprises entre 2 DS de part et d'autre de la moyenne et la distribution serait symétrique autour de la moyenne. D'après les données de cet exercice, il faudrait que les niveaux de nicotine puissent être négatifs or, ce n'est pas possible.

	Douleur et raideur	Pas de douleur ni de raideur	Total
Poignets hypermobiles	5	91	96
Poignets normaux	102	462	564
Total	107	553	192

chiffre 5 dans la cellule supérieure gauche. Le reste des cellules du tableau se remplit de la même façon.

Calculer un risque relatif dans les études transversales a du sens. Le risque relatif vaut 0,29. Le risque d'avoir des poignets raides ou douloureux pour des musiciens qui ont des poignets hypermobiles vaut seulement 29 % du risque des musiciens qui n'ont pas de poignet hypermobile. Au moyen d'Instat ou de l'équation 8.7, vous pouvez calculer l'IC approximatif à 95 %. Il s'étend de 0,12 à 0,69. Nous sommes sûrs à 95 % que les musiciens aux articulations hypermobiles ont une prévalence de douleur et de raideur des poignets qui vaut entre 12 % et 69 % de celle des musiciens aux poignets normaux. Les poignets hypermobiles protègent donc les musiciens de la douleur et de la raideur.

Vous pourriez également calculer la différence entre les prévalences. Parmi les musiciens aux poignets hypermobiles, la prévalence de douleur et de souffrance vaut 2,5 %. Parmi les musiciens sans poignets hypermobiles, la prévalence de douleur et de souffrance vaut 18,1 %. La différence est de 12,9 %, avec un IC approximatif à 95 % de la différence allant de 7,4 % à 18,4 %.

Pour interpréter ces résultats, vous devez supposer que les musiciens étudiés dans cette étude sont représentatifs des musiciens d'autres lieux. Vous devez aussi supposer que l'évaluation de l'hypermobilité est objective et qu'elle n'est pas influencée par la présence ou l'absence de douleur et de raideur.

3. Un même nombre de cellules (100 000 par ml) a été placé dans quatre flacons. Deux lignées de cellules ont été utilisées. Certains flacons ont été traités avec des substances pharmaceutiques, tandis que les autres n'ont été traités qu'avec l'adjuvant (contrôle). Les données reprises dans le tableau suivant correspondent aux nombres moyens de cellules (milliers par millilitre) après 24 heures. Analysez ces données de la façon la plus complète possible. Si vous aviez accès à toutes les données originales, souhaiteriez-vous résumer les données autrement ? Comment ?

   Il ne s'agit *pas* d'un tableau de contingence et aucune des méthodes présentées dans le chapitre 8 n'est appropriée. Les valeurs reprises dans le tableau sont des nombres moyens de cellules et pas des nombres de sujets dans chaque catégorie.

	Substance pharmaceutique	Pas de substance pharmaceutique
Lignée de cellules 1	145	198
Lignée de cellules 2	256	356

données sont impressionnantes et apportent une solide évidence que le système de distribution d'eau peut être impliqué dans l'étiologie de la maladie.

Pour interpréter ces données, vous devez supposer que les sujets et les témoins étudiés sont représentatifs des autres personnes qui ont ou qui n'ont pas contracté la maladie. Vous devez également supposer que les sujets et les témoins ont été sélectionnés indépendamment. Par exemple, l'étude ne serait pas valide si les chercheurs avaient sélectionné plusieurs cas ou témoins du même ménage (ils boivent bien sûr tous de l'eau du même système de distribution). Et enfin, vous devez supposer qu'il n'y a pas d'autres différences entre les personnes qui boivent aux deux systèmes de distribution. Vous ne seriez pas capables d'interpréter ces résultats si ces deux groupes de personnes achetaient aussi leur nourriture à deux endroits différents ou travaillaient dans des villages différents ou différaient d'une quelque autre façon.

Comment ces chercheurs auraient-ils pu mener une étude prospective pour tester leur hypothèse ?

Pour mener une étude prospective, les chercheurs auraient dû sélectionner des sujets sur base du système de distribution qui les approvisionnait en eau de boisson et observer ensuite s'ils développaient la maladie. Si la maladie est rare, une étude prospective exige plus de sujets qu'une étude rétrospective.

9 CHAPITRE 10

1. Vous souhaitez tester l'hypothèse selon laquelle un lancé de pièce n'est pas truqué. Vous lancez 6 fois la pièce et elle tombe chaque fois sur le côté face. Quelle est l'hypothèse nulle ? Que vaut la P-valeur ? Quel test est le plus approprié, un test uni ou bilatéral ? Que tirez-vous comme conclusion ?

   L'hypothèse nulle est que la pièce est lancée honnêtement et donc que la probabilité qu'elle tombe sur le côté face soit donc exactement 0,5. Si c'était vrai, la probabilité que 6 lancés consécutifs tombent tous sur le côté face serait $0{,}5^6 = 1/64 = 0{,}016$. Il s'agit d'une P-valeur unilatérale. La probabilité que le résultat des 6 lancés soient identiques (tous « pile » ou tous « face ») vaut deux fois 0,016 soit 3,2 %. C'est une P-valeur bilatérale.

   Il est impossible d'interpréter ces résultats sans la moindre information sur le contexte. Pour quelle raison pourrait-on suspecter que le lancé de la pièce soit truqué ? Si vous utilisez une P-valeur unilatérale, vous deviez avoir décidé, avant de collecter les données, que vous vous attendiez à observer trop de côtés face. Dans le cas contraire, vous devriez choisir une P-valeur bilatérale de 0,032. Comment doit-on interpréter les résultats ? Soit il existe un problème systématique avec le lancé de la pièce qui a pour conséquence que le côté face apparaît plus de la moitié des fois, soit il s'agit d'une coïncidence. Cette coïncidence surviendrait avec une pièce non truquée 3,2 % des fois. Sauf si on a une raison de suspecter que le lancé de la pièce est truqué (par exemple, s'il faisait partie d'un spectacle de prestidigitation), on aurait plutôt tendance à penser qu'il s'agit d'une coïncidence.

des enfants non daltoniens qui peuvent répondre correctement à une question concernant la couleur à cet âge.

4. Pour le patient C de l'exemple de la porphyrie, quelle serait la valeur prédictive d'un test positif si, par votre intuition clinique, vous estimiez la probabilité a priori à 75 % ?

Vous pouvez construire le tableau suivant.

	Présence de la maladie	Absence de la maladie	Total
< 99 unités	615	9	624
> 99 unités	135	241	376
Total	750	250	1000

La valeur prédictive vaut 615/624 = 98,6 %. Puisque la probabilité a priori est plus grande, la valeur prédictive doit être plus grande également. Avant d'obtenir les résultats du test, vous pensiez déjà que la probabilité serait plus grande. Les résultats du test augmentent encore davantage la probabilité.

13 CHAPITRE 15

1. Une étudiante voudrait savoir si le traitement de cellules avec une hormone spécifique augmente le nombre d'un certain type de récepteurs. En accord avec son promoteur, elle estime qu'une augmentation de moins de 100 récepteurs par cellule est trop faible pour y prêter attention. Sur base de la déviation standard de résultats observés dans des études similaires, elle calcule la taille de l'échantillon nécessaire pour avoir une puissance de 90 % pour détecter une croissance de 100 récepteurs par cellule. Elle effectue l'expérience autant de fois que nécessaire, rassemble les données et obtient une P-valeur de 0,04.

L'étudiante pense que l'expérience est pertinente et elle pensait que la probabilité a priori que son hypothèse soit vraie était de 60 %. Son promoteur était plus sceptique et pensait que la probabilité a priori atteignait seulement 5 %.

a) En combinant la probabilité a priori et la P-valeur, quelle est la probabilité que ces résultats soient dus au hasard ? Répondez à la fois du point de vue de l'étudiant et du point de vue du promoteur.

Le point de vue de l'étudiante est illustré dans le tableau suivant. Il présente les résultats de 1000 expériences hypothétiques.

Parmi 556 expériences ayant une P-valeur inférieure à 0,04, le nombre de récepteurs augmente réellement dans 540. La probabilité que le nombre de récepteurs augmente réellement vaut donc 540/556 = 97,1 % et il reste 2,9 % de chance que les résultats soient dus à une coïncidence.

Le point de vue du promoteur, illustré dans le tableau suivant, est différent.

tistiquement significatif si l'hypothèse de recherche est vraie que si l'hypothèse nulle est vraie.

Vous rencontrerez rarement, voire jamais, un rapport de vraisemblance calculé pour un test statistique. C'est en partie parce que, dans de nombreuses situations, il est difficile de définir l'hypothèse de recherche et donc difficile de calculer la puissance. Mais c'est aussi simplement par habitude.

15 CHAPITRE 17

1. Dans l'exemple 17.1, comment les chercheurs auraient-ils analysé les données s'ils avaient mesuré deux fois la sensibilité à l'insuline et le pourcentage de C20-22 chez chaque sujet ?

   L'approche la plus directe consisterait à représenter chaque paire de mesures par un point sur le graphique. Si faites cela, vous aurez n = 26 au lieu de n = 13. Mais il n'est pas correct d'analyser les données puisque vous ne respecteriez pas l'hypothèse d'indépendance des observations. Il est très probable que la seconde mesure de chaque sujet soit plus proche de sa propre première mesure que des mesures des autres sujets.

   Vous devriez d'abord obtenir la moyenne des deux mesures de sensibilité à l'insuline pour chaque sujet, ensuite la moyenne des deux mesures de %C20-22 de chaque sujet et enfin calculer le coefficient de corrélation entre la mesure moyenne de sensibilité à l'insuline et la mesure moyenne de %C20-22.

2. La P-valeur de l'exemple 17.1 était bilatérale. Que vaut la P-valeur unilatérale ? Que signifie-t-elle ?

   La P-valeur bilatérale vaut 0,0021. Rappelez-vous ce que cela signifie. S'il n'y avait vraiment pas de corrélation entre les variables X et Y dans la population, on aurait 2,1 % de chance d'observer une corrélation aussi forte (ou plus forte) dans une étude de cette taille. Si l'hypothèse nulle était vraie, on aurait 0,1 % de chance de trouver un r = 0,77 en tirant 13 sujets au hasard et 0,1 % de chance de trouver un r = 0,77 en tirant 13 sujets au hasard.

   Vous ne pouvez calculer une P-valeur unilatérale que si vous l'aviez décidé avant de collecter les données et que vous aviez spécifié la direction de l'hypothèse de recherche. Si vous aviez énoncé que la corrélation serait positive (la sensibilité à l'insuline augmente quand le %C20-22 augmente), alors la P-valeur unilatérale serait égale à 0,001.

3. Faut-il mesurer les variables X et Y dans les mêmes unités pour calculer un coefficient de corrélation ? Peuvent-elles être mesurées dans les mêmes unités ?

   Comme le montre l'exemple 17.1, X et Y ne doivent pas être mesurés dans les mêmes unités. Mais, ils peuvent être mesurés dans les mêmes unités.

4. Que vaut la P-valeur si r = 0,5 avec un échantillon de taille n = 10 ? Que vaut la P-valeur si r = 0,5 avec un échantillon de taille n = 100 ?

b) Combien de sujets sont nécessaires pour avoir 60 % de puissance de détecter une diminution de 50 % du taux de mortalité ($\alpha = 0,1$) ?

$p_1 = 0,10$, $p_2 = 0,5 \cdot p_1 = 0,05$, $\alpha = 0,10$, $\beta = 0,40$

Vous ne pouvez pas utiliser le tableau du chapitre pour trouver la somme des carrés des valeurs z puisque la table n'inclut pas $\beta = 0,40$. Regardez la deuxième colonne du tableau A4.2 de l'annexe pour trouver 40 %, vous pouvez ensuite constater que z_β vaut environ 0,25 (40 % d'une population gaussienne à $z > 0,25$). Pour $\alpha = 0,10$ (bilatérale), il faut utiliser la quatrième colonne du tableau A4.2 pour trouver que z vaut environ 1,65 (10 % d'une population gaussienne a $z > 1,65$ ou $z < 1,65$).

Maintenant, vous pouvez remplacer toutes les inconnues dans l'équation 22.6. Vous avez besoin de 200 personnes dans chaque groupe. Pour finir, vous obtenez un nombre raisonnable.

c) Est-il éthique de comparer une nouvelle technique (que vous pensez être meilleure) avec une technique standard ?

Si vous êtes vraiment certains que la nouvelle technique est supérieure, il n'est pas éthique d'appliquer l'ancienne technique à qui que ce soit. Mais à quel point êtes-vous vraiment sûrs ? L'histoire de la médecine est remplie d'exemples de nouvelles techniques qui se sont avérées sans intérêt malgré l'enthousiasme des personnes qui les avaient utilisées dès le début. Pour vraiment comparer deux techniques, il faut réaliser une étude. Il se peut que les patients qui ont reçu le traitement standard se rétablissent mieux que les patients qui ont reçu le nouveau traitement.

6. Les lymphocytes contiennent des récepteurs bêta-adrénergiques. L'épinéphrine se lie à ces récepteurs et module leurs réponses immunitaires. Il est possible de compter le nombre moyen de récepteurs dans les lymphocytes humains en utilisant un petit échantillon de sang. Vous souhaitez tester l'hypothèse que les personnes asthmatiques ont moins de récepteurs. En lisant des articles, vous apprenez qu'il y a environ 1000 récepteurs par cellule et que le coefficient de variation dans une population normale est d'environ 25 %.

a) De combien de sujets asthmatiques avez-vous besoin pour déterminer le nombre de récepteurs par cellule, à plus ou moins 10 récepteurs près, avec 90 % de confiance ?

Vous devez utiliser l'équation 22.1. Puisque vous savez que le coefficient de variation vaut 25 %, vous pouvez déduire que la déviation standard vaut environ 250 récepteurs/cellule. L'équation 22.1 permet de calculer la taille d'échantillon pour des intervalles confiance à 95 %. Pour des intervalles à 90 % (qui sont plus étroits), la valeur de t^* doit être remplacée par un nombre plus petit. La table A5.3 de l'annexe vous montre que la valeur de t^* pour 90 % de confiance et de nombreux degrés de liberté vaut 1,65.

Donc $n = 1,65^2 \times (250/100)^2 = 17$ personnes dans chaque groupe.

b) Vous voulez comparer un groupe de sujets normaux avec un groupe d'asthmatiques. Combien de sujets sont nécessaires dans chaque groupe pour avoir

Si vous souhaitez analyser les données de façon manuelle, la première étape consiste à ordonner les valeurs sans distinguer le fait qu'il s'agisse de témoins ou de sujets traités avec le médicament. Il faut ensuite additionner les rangs dans chaque groupe.

Témoins	Rang	Médicament	Rang
1162	8	892	2
1095	6	903	3
1327	12	1164	9
1261	11	1002	5
1103	7	961	4
1235	10	875	1
Somme des rangs	54	Somme des rangs	24

De nombreux livres (pas celui-ci) fournissent des tables pour le test de Mann-Whitney. On peut y trouver la P-valeur sur base de la somme des rangs et de la taille de l'échantillon.

Pour la calculer manuellement, il faut utiliser les équations 24.2 et 24.3, qui donnent des réponses approximatives. L'équation 24.2 permet de calculer deux valeurs pour U. $U = 33$ et $U = 3$. Vous obtiendrez la même valeur de Z quelle que soit la valeur de U utilisée. Sur base de l'équation 24.3 (et sachant que $n_a = n_b = 6$), $z = 2,40$. On trouve dans la table A5.2 de l'annexe que pour 1,64 % d'une distribution gaussienne, $z > 2,40$ ou $< -2,40$. La P-valeur bilatérale vaut donc approximativement 0,0164.

b) Calculez le test t.

Une fois de plus, il serait plus facile d'utiliser un programme informatique. Instat donne les résultats suivants : $t = 3,978$ et $P = 0,0026$.

Pour calculer le test t manuellement, vous devez d'abord calculer la moyenne et l'ESM de chaque groupe.

Témoins :	Moyenne = 1197,2	ESM = 37,8
Traités :	Moyenne = 966,2	ESM = 44,1

Vous devez ensuite calculer l'erreur standard de la différence entre les moyennes qui est égale à la racine carrée de la somme du carré des valeurs des deux ESM. Cela vaut 58,1. Le rapport t est égal à la différence entre les moyennes divisée par l'erreur standard de la différence = $(1197,2 - 96,2)/58,1 = 3,97$. Pour déterminer le nombre de degrés de liberté, il faut additionner le nombre total de sujets et soustraire 2. dl = 10. Regardez dans la table A5.7 de l'annexe : dl = 10 et P = 0,003.

facteur 3 ou avec une augmentation du risque par un facteur 3. Même si l'étude ne montre aucune évidence d'association, elle est trop petite pour être négative de façon convaincante.

c) Si les implants doublaient réellement le risque de maladie du tissu conjonctif, quelle est la puissance d'une étude de cette taille pour détecter une différence significative avec $P < 0,05$?

Une façon de juger de l'utilité des données de l'étude est de calculer la puissance qu'avait l'étude pour mettre en évidence des différences hypothétiques. Par exemple, si les femmes avec des implants mammaires développaient réellement deux fois plus de maladies du tissu conjonctif que les non exposées, quelle était la puissance de cette étude pour trouver une différence statistiquement significative ?

Nous fixerons p_1 à 0,0067, puisqu'il s'agit de la proportion de non exposées (et d'exposées) qui ont développé une maladie du tissu conjonctif (10/1484). Nous fixerons p_2 à la moitié de cette valeur, soit 0,0034. Pour calculer la puissance, vous devez d'abord utiliser l'équation 27.3 pour calculer que $H = 0,0472$.

La figure 27.1 n'est pas très utile dans le cas de si petites valeurs de H, vous devez donc utiliser l'équation 27.4. Fixez $z_a = 1,96$ puisque nous utilisons la définition conventionnelle de la signification statistique ($P < 0,05$, bilatérale). Puisque les tailles d'échantillon sont différentes, vous devez calculer la moyenne harmonique selon l'équation 27.5. $n = 2 \times 747 \times 1484/(747 + 1484) = 993,7$. Maintenant, utilisez l'équation 27.4 pour calculer $z_{puissance} = 0,91$. La table A4.14 de l'annexe vous apprend que la puissance vaut environ 18 %. Si les implants mammaires doublaient réellement le risque de maladie du tissu conjonctif, une étude de cette taille permettrait d'observer une association significative moins d'une fois sur cinq. Puisque cette étude a une puissance si basse, vous ne devez pas interpréter les résultats trop fermement.

23 CHAPITRE 28

1. Vous utilisez un hémocytomètre pour dénombrer des globules blancs. Vous regardez une lame quadrillée au microscope et observez 25 carrés de 0,1 microlitre. Vous dénombrez les globules blancs dans 9 carrés et vous en trouvez 50 au total. Pouvez-vous calculer l'intervalle de confiance à 95 % du nombre de globules blancs par microlitre ? Quelles hypothèses faites-vous ?

   Vous avez compté 50 cellules dans 9 carrés. Puisqu'il y a 250 carrés par microlitre, il y a $50 \times (250/9) = 1389$ cellules/microlitre. Pour calculer l'intervalle de confiance, il faut travailler avec le nombre réellement compté, 50. Vous pouvez trouver dans la table A5.9 de l'annexe que l'intervalle de confiance à 95 % du nombre de cellules dans 9 carrés s'étend de 37,1 à 65,9. Il faut multiplier chaque limite de confiance par (250/9) pour déterminer l'intervalle de confiance à 95 % du

TABLE A5.1 Intervalle de confiance à 95 % d'une proportion

Il faut repérer la colonne correspondant au numérateur (N) et la ligne correspondant au dénominateur (D). Les valeurs dans la table sont les limites inférieures et supérieures de l'intervalle de confiance à 95 % de la proportion. Par exemple, si on constate que 4 patients sur 20 présentent une certaine complication, la proportion vaut 0,200 et l'IC à 95 % va de 0,057 à 0,437.

D	Numérateur																			
	0		1		2		3		4		5		6		7		8		9	
1	0,000	0,975	0,025	1,00																
2	0,000	0,842	0,013	0,987	0,158	1,00														
3	0,000	0,708	0,008	0,906	0,094	0,992	0,292	1,00												
4	0,000	0,602	0,006	0,806	0,068	0,932	0,194	0,994	0,398	1,00										
5	0,000	0,522	0,005	0,716	0,053	0,853	0,147	0,947	0,284	0,995	0,478	1,00								
6	0,000	0,459	0,004	0,641	0,043	0,777	0,118	0,882	0,223	0,957	0,359	0,996	0,541	1,00						
7	0,000	0,410	0,004	0,579	0,037	0,710	0,099	0,816	0,184	0,901	0,290	0,963	0,421	0,996	0,590	1,00				
8	0,000	0,369	0,003	0,527	0,032	0,651	0,085	0,755	0,157	0,843	0,245	0,915	0,349	0,968	0,473	0,997	0,631	1,00		
9	0,000	0,336	0,003	0,482	0,028	0,600	0,075	0,701	0,137	0,788	0,212	0,863	0,299	0,925	0,400	0,972	0,518	0,997	0,664	1,00
10	0,000	0,308	0,003	0,445	0,025	0,556	0,067	0,652	0,122	0,738	0,187	0,813	0,262	0,878	0,348	0,933	0,444	0,975	0,555	0,997
11	0,000	0,285	0,002	0,413	0,023	0,518	0,060	0,610	0,109	0,692	0,167	0,766	0,234	0,833	0,308	0,891	0,390	0,940	0,482	0,977
12	0,000	0,265	0,002	0,385	0,021	0,484	0,055	0,572	0,099	0,651	0,152	0,723	0,211	0,789	0,277	0,848	0,349	0,901	0,428	0,945
13	0,000	0,247	0,002	0,360	0,019	0,454	0,050	0,538	0,091	0,614	0,139	0,684	0,192	0,749	0,251	0,808	0,316	0,861	0,386	0,909
14	0,000	0,232	0,002	0,339	0,018	0,428	0,047	0,508	0,084	0,581	0,128	0,649	0,177	0,711	0,230	0,770	0,289	0,823	0,351	0,872
15	0,000	0,218	0,002	0,319	0,017	0,405	0,043	0,481	0,078	0,551	0,118	0,616	0,163	0,677	0,213	0,734	0,266	0,787	0,323	0,837
16	0,000	0,206	0,002	0,302	0,016	0,383	0,040	0,456	0,073	0,524	0,110	0,587	0,152	0,646	0,198	0,701	0,247	0,753	0,299	0,802
17	0,000	0,195	0,001	0,287	0,015	0,364	0,038	0,434	0,068	0,499	0,103	0,560	0,142	0,617	0,184	0,671	0,230	0,722	0,278	0,770
18	0,000	0,185	0,001	0,273	0,014	0,347	0,036	0,414	0,064	0,476	0,097	0,535	0,133	0,590	0,173	0,643	0,215	0,692	0,260	0,740
19	0,000	0,176	0,001	0,260	0,013	0,331	0,034	0,396	0,061	0,456	0,091	0,512	0,126	0,566	0,163	0,616	0,203	0,665	0,244	0,711
20	0,000	0,168	0,001	0,249	0,012	0,317	0,032	0,379	0,057	0,437	0,087	0,491	0,119	0,543	0,154	0,592	0,191	0,639	0,231	0,685
21	0,000	0,161	0,001	0,238	0,012	0,304	0,030	0,363	0,054	0,419	0,082	0,472	0,113	0,522	0,146	0,570	0,181	0,616	0,218	0,660
22	0,000	0,154	0,001	0,228	0,011	0,292	0,029	0,349	0,052	0,403	0,078	0,454	0,107	0,502	0,139	0,549	0,172	0,593	0,207	0,636
23	0,000	0,148	0,001	0,219	0,011	0,280	0,028	0,336	0,050	0,388	0,075	0,437	0,102	0,484	0,132	0,529	0,164	0,573	0,197	0,615
24	0,000	0,142	0,001	0,211	0,010	0,270	0,027	0,324	0,047	0,374	0,071	0,422	0,098	0,467	0,126	0,511	0,156	0,553	0,188	0,594
25	0,000	0,137	0,001	0,204	0,010	0,260	0,025	0,312	0,045	0,361	0,068	0,407	0,094	0,451	0,121	0,494	0,149	0,535	0,180	0,575
26	0,000	0,132	0,001	0,196	0,009	0,251	0,024	0,302	0,044	0,349	0,066	0,394	0,090	0,436	0,116	0,478	0,143	0,518	0,172	0,557
27	0,000	0,128	0,001	0,190	0,009	0,243	0,024	0,292	0,042	0,337	0,063	0,381	0,086	0,423	0,111	0,463	0,138	0,502	0,165	0,540
28	0,000	0,123	0,001	0,183	0,009	0,235	0,023	0,282	0,040	0,327	0,061	0,369	0,083	0,410	0,107	0,449	0,132	0,487	0,159	0,524
29	0,000	0,119	0,001	0,178	0,008	0,228	0,022	0,274	0,039	0,317	0,058	0,358	0,080	0,397	0,103	0,435	0,127	0,472	0,153	0,508
30	0,000	0,116	0,001	0,172	0,008	0,221	0,021	0,265	0,038	0,307	0,056	0,347	0,077	0,386	0,099	0,423	0,123	0,459	0,147	0,494

TABLE A5.3 Valeurs de t* à utiliser pour calculer les intervalles de confiance

Cette table fournit les valeurs de t* nécessaires pour calculer les intervalles de confiance. L'intervalle de confiance de la moyenne s'étend dans les deux directions de part et d'autre de la moyenne, sur une distance égale à l'ESM multipliée par la valeur critique de t présentée dans la table ci-dessous. Le nombre de degrés de liberté est égal au nombre d'observations moins un. Dans ce livre, le symbole t* est utilisé pour désigner la valeur critique de la distribution t.

	Valeur critique de t pour l'IC à _ %		
dl	IC à 90 %	IC à 95 %	IC à 99 %
1	6,3137	12,7062	63,6559
2	2,9200	4,3027	9,9250
3	2,3534	3,1824	5,8408
4	2,1318	2,7765	4,6041
5	2,0150	2,5706	4,0321
6	1,9432	2,4469	3,7074
7	1,8946	2,3646	3,4995
8	1,8585	2,3060	3,3554
9	1,8331	2,2622	3,2498
10	1,8125	2,2281	3,1693
11	1,7959	2,2010	3,1058
12	1,7823	2,1788	3,0545
13	1,7709	2,1604	3,0123
14	1,7613	2,1448	2,9768
15	1,7531	2,1315	2,9467
16	1,7459	2,1199	2,9208
17	1,7396	2,1098	2,8982
18	1,7341	2,1009	2,8784
19	1,7291	2,0930	2,8609
20	1,7247	2,0860	2,8453
21	1,7207	2,0796	2,8314
22	1,7171	2,0739	2,8188
23	1,7139	2,0687	2,8073
24	1,7109	2,0639	2,7970
25	1,7081	2,0595	2,7874
26	1,7056	2,0555	2,7787
27	1,7033	2,0518	2,7707
28	1,7011	2,0484	2,7633
29	1,6991	2,0452	2,7564
30	1,6973	2,0423	2,7500
35	1,6896	2,0301	2,7238
40	1,6839	2,0211	2,7045
45	1,6794	2,0141	2,6896
50	1,6759	2,0086	2,6778
100	1,6602	1,9840	2,6259
200	1,6525	1,9719	2,6006
500	1,6479	1,9647	2,5857
1000	1,6464	1,9623	2,5807

TABLE A5.9 Intervalles de confiance à 95 % pour la distribution de Poisson

C est le nombre d'événements par unité de temps ou le nombre d'objets par unité d'espace. La table fournit l'intervalle de confiance à 95 % du nombre moyen d'événements par unité de temps ou du nombre moyen d'objets par unité d'espace. Pour des valeurs de C plus grandes, il faut utiliser l'équation suivante : $C - 1{,}96\sqrt{C}$ à $C + 1{,}96\sqrt{C}$. Il ne faut pas normaliser C et l'exprimer en une unité standard quelconque, C est le nombre d'événements ou d'objets.

C	Inférieure	Supérieure	C	Inférieure	Supérieure	C	Inférieure	Supérieure
0	0	3,69	41	29,42	55,62	82	65,21	101,78
1	0,03	5,57	42	30,27	56,77	83	66,11	102,89
2	0,24	7,22	43	31,12	57,92	84	67	104
3	0,62	8,77	44	31,97	59,07	85	67,89	105,10
4	1,09	10,24	45	32,82	60,21	86	68,79	106,21
5	1,62	11,67	46	33,68	61,36	87	69,68	107,31
6	2,20	13,06	47	34,53	62,50	88	70,58	108,42
7	2,81	14,42	48	35,39	63,64	89	71,47	109,52
8	3,45	15,76	49	36,25	64,78	90	72,37	110,63
9	4,12	17,08	50	37,11	65,92	91	73,27	111,73
10	4,80	18,39	51	37,97	67,06	92	74,16	112,83
11	5,49	19,68	52	38,84	68,19	93	75,06	113,93
12	6,20	20,96	53	39,70	69,33	94	75,96	115,03
13	6,92	22,23	54	40,57	70,46	95	76,86	116,13
14	7,65	23,49	55	41,43	71,59	96	77,76	117,23
15	8,40	24,74	56	42,30	72,72	97	78,66	118,33
16	9,15	25,98	57	43,17	73,85	98	79,56	119,43
17	9,90	27,22	58	44,04	74,98	99	80,46	120,53
18	10,67	28,45	59	44,91	76,11	100	81,36	121,63
19	11,44	29,67	60	45,79	77,23	101	82,27	122,72
20	12,22	30,89	61	46,66	78,36	102	83,17	123,82
21	13	32,10	62	47,54	79,48	103	84,07	124,92
22	13,79	33,31	63	48,41	80,60	104	84,98	126,01
23	14,58	34,51	64	49,29	81,73	105	85,88	127,11
24	15,38	35,71	65	50,17	82,85	106	86,78	128,20
25	16,18	36,90	66	51,04	83,97	107	87,69	129,30
26	16,98	38,10	67	51,92	85,09	108	88,59	130,39
27	17,79	39,28	68	52,80	86,21	109	89,50	131,49
28	18,61	40,47	69	53,69	87,32	110	90,41	132,58
29	19,42	41,65	70	54,57	88,44	111	91,31	133,67
30	20,24	42,83	71	55,45	89,56	112	92,22	134,77
31	21,06	44	72	56,34	90,67	113	93,13	135,86
32	21,89	45,17	73	57,22	91,79	114	94,04	136,95
33	22,72	46,34	74	58,11	92,90	115	94,94	138,04
34	23,55	47,51	75	58,99	94,01	116	95,85	139,13
35	24,38	48,68	76	59,88	95,13	117	96,76	140,22
36	25,21	49,84	77	60,77	96,24	118	97,67	141,31
37	26,05	51	78	61,66	97,35	119	98,58	142,40
38	26,89	52,16	79	62,55	98,46	120	99,49	143,49
39	27,73	53,31	80	63,44	99,57	121	100,40	144,58
40	28,58	54,47	81	64,33	100,68	122	101,31	145,67

TABLE DES MATIÈRES

PARTIE I

CHAPITRE 12

CHAPITRE 13



Biostatistique

Cet ouvrage présente les **principales méthodes de statistique médicale** utilisées dans la pratique quotidienne et dans la littérature théorique.

Par une approche claire et synthétique, il aborde de manière détaillée les concepts et les techniques de base tels qu'intervalle de confiance, test d'hypothèse, P-valeur, test t, test χ^2, corrélation et régression. Il introduit également les tests non paramétriques, la survie, l'analyse de variance à un facteur et des techniques plus poussées telles que la régression linéaire multiple, logistique et non linéaire.

L'auteur insiste tout particulièrement sur **l'interprétation des résultats**, les **conditions et situations d'application des méthodes** et il met l'accent sur les **aspects pratiques**. Les sections plus techniques ou avancées sont signalées et peuvent être sautées sans conséquence pour la compréhension de l'ensemble. Des **exemples médicaux** illustrent la théorie et des exercices résolus en détails figurent en fin de chaque chapitre.

HARVEY J. MOTULSKY

Il enseigne la pharmacologie et la biostatistique à l'Université de Californie à San Diego. Il est également président-fondateur de GraphPad Software, Inc.

MOTULSKY	110092
ISBN	2-7445-0092-5
ISSN	1376-2281

